Weighted Sum Rate Maximization in Full-Duplex Multi-User Multi-Cell MIMO Networks

Weighted Sum Rate Maximization in Full-Duplex Multi-User Multi-Cell MIMO Networks

Paula Aquilina, *Student Member, IEEE*, Ali Cagatay Cirik, *Member, IEEE*, and Tharmalingam Ratnarajah, *Senior Member, IEEE*

***Abstract*—In this paper we focus on a multi-user multi-cell scenario with full-duplex (FD) base-stations (BSs) and half-duplex (HD) downlink (DL) and uplink (UL) users, where all nodes are equipped with multiple antennas. Our goal is to design filters for weighted sum rate (WSR) maximization whilst taking into consideration the effect of transmitter and receiver distortion. Since WSR problems are non-convex we exploit the relationship between rate and mean squared error (MSE) in order to propose low complexity alternating optimization algorithms which are guaranteed to converge. While the initial design assumes perfect channel state information (CSI), we also move beyond this assumption and consider WSR problems under imperfect CSI. This is done using two types of error models; the first is a norm-bounded error model, suitable for cases where the CSI error is dominated by quantization issues, and the second is a stochastic error model, suitable for errors that occur during the channel estimation process itself. Results show that rates achieved in FD mode are higher than those achieved by the baseline HD schemes and demonstrate the robust performance of the proposed imperfect CSI designs. Additionally we also extend our original WSR problem to one which maximizes the total DL rate subject to each UL user achieving a desired target rate. This latter design can be used to overcome potential unfairness issues and ensure that all UL users are equally served in every time slot.**

***Index Terms*—Filter design, full-duplex, mean squared error, MIMO, multi-cell, weighted sum rate maximization.**

I. Introduction

THE demand for mobile wireless network resources is constantly on the rise, pushing for new communication technologies that are able to support unprecedented rates. One such contender is full-duplex (FD) communication. Whilst traditional half-duplex (HD) systems require separate time or frequency resources for downlink (DL) and uplink (UL) communication, FD considers simultaneous DL and UL transmission. The potential to significantly improve spectral efficiency makes FD communication an attractive candidate solution to the ever growing spectrum demand problem.

Even though the possible benefits of FD operation are easy to foresee, there are implementation issues that may pose significant challenges when trying to translate the theoretical gains into practical ones. Self-interference (SI), where power from the DL transmission interferes with the UL received signal at FD nodes has traditionally been considered a major stumbling block. In fact FD communication was conventionally believed to be infeasible due to the presence of SI. However, in recent years, there have been a number of breakthroughs in hardware design showing that SI can be canceled up to acceptable levels [1]–[5] and demonstrating the feasibility of FD nodes.

The promise of increased spectral efficiency, alongside with the newfound ability to mitigate SI, has motivated a wide range of research into FD communication and its possible applications. For example, the use of FD operation in relays [6], [7] and cognitive radio systems [8], [9] has proven to be effective. Additionally FD operation, either at the base-station (BS) only [10] or at both the users and the BS [11], has been found to be particularly suited for small cell scenarios due to the low transmit powers and small transmission distances involved. Different to [6]–[11], which consider single-cell systems, here we focus on a more practical multi-cell system with FD BSs and HD users; the multi-cell aspect introduces the additional challenge of co-channel-interference (CCI) from nodes in other cells. A similar system has been considered in [12] where the authors focus on user selection and power allocation methods. A stochastic geometry approach for system performance characterization of FD multi-cell systems has been considered in [13]–[15]. In contrast to [12]–[15], which assume all nodes are equipped with a single antenna, we consider a multiple-input multiple-output (MIMO) system and focus on beamformer design for weighted sum rate (WSR) maximization.

As was hinted earlier, in this work we focus on a multi-cell scenario where each BS serves multiple HD users; however unlike traditional systems, the BSs operate in FD mode serving all of their corresponding DL and UL users simultaneously. The FD capability at the BSs and the inherent structure of the network lead to a large amount of interference at the different receivers. Fig. 1 provides a simple illustration of the network under consideration, having G cells and one DL and one UL user per cell. It can be seen that, apart from the usual standard HD network interference components, for UL communication BSs have additional SI and BS-to-BS interference, while DL users have additional CCI from UL users both from the same cell and from other cells.

Since our main focus is on small cell networks where coverage distances are short and BSs and users have similar transmission powers [16], we consider the case where none of the interference components may be ignored; therefore

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/L025299/1.

P. Aquilina and T. Ratnarajah are with the Institute of Digital Communications, School of Engineering, The University of Edinburgh, Alexander Graham Bell Building, Kings Buildings, Mayfield Road, Edinburgh, EH9 3JL, United Kingdom (email: {p.aquilina, t.ratnarajah}@ed.ac.uk).

A.C. Cirik is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. (email: cirik@ece.ubc.ca)

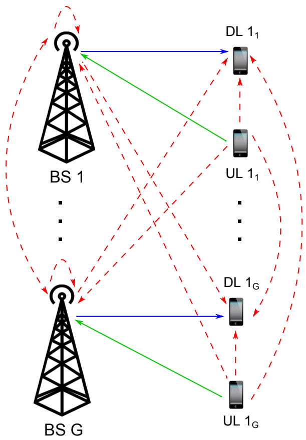


Fig. 1: G-cell network with an FD BS, and one DL and one UL user per cell. Solid lines represent desired links, while dashed ones represent interference links.

the challenge is to manage all the interference while still delivering good service to all users. This is in contrast to prior studies which assume that CCI can be avoided via scheduling [17], allocating different sub-carriers [18] or assuming channels between UL and DL users to be sufficiently weak [19]. Additionally, we also take into consideration the effect of transmitter and receiver distortion. These hardware impairments are a natural consequence of non-ideal amplifiers, oscillators, analog-to-digital converters (ADCs) and digital-to-analog converters (DACs), and cannot be avoided in practice [7], [9], [20].

Within this context, our aim is to investigate under what conditions replacing HD BSs with FD ones may be beneficial within a small cell scenario with multiple cells and multiple legacy HD users. Since WSR problems are non-convex, we map each of them to a weighted minimum mean squared error (WMMSE) problem. This technique is less computationally complex than gradient-based alternatives for WSR maximization, is guaranteed to converge, and has been proven to work for various types of HD networks [21]–[23]. The rate to MSE relationship was also used for transceiver design in MIMO interference channels with FD nodes throughout under the perfect CSI assumption in [20]. Unlike [20], we consider HD users and cater for multiple users per cell; additionally, we cater for imperfect CSI under two different models. To the best of our knowledge, the potential gains of FD operation in a multi-cell multi-user MIMO system have not been reported yet. Motivated by this, here we make an attempt to understand the benefits and actual gains that can be achieved by the use of FD-based transceivers in such systems.

With respect to the imperfect CSI aspect, we first consider a norm-bounded error model, suitable for situations where the CSI error is mainly due to quantization. Secondly we consider a stochastic CSI error model, more suited to errors occurring during the channel estimation process itself. Results show that FD communication can indeed achieve higher rates than the baseline HD scheme for intermediate to low distortion levels and confirm the robust performance of the imperfect CSI designs. Finally, we extend our original design to one which maximizes the total DL rate subject to each UL user achieving a pre-established target rate. This can be used in situations where it is important that each UL user is equally served in every time slot, which is not guaranteed with the joint design.

The rest of the paper is organized as follows. Section II provides some preliminaries. In Section III we present the WSR problem under perfect CSI. Sections IV and V tackle the norm-bounded error and the stochastic error problems respectively. Next in Section VI we consider the extension to a weighted DL rate maximization problem subject to a minimum per UL user target rate. Simulation results are presented in Section VII. Section VIII provides an insight on the implementation and complexity of the proposed algorithms, and finally Section IX presents some concluding remarks.

Notation: $|\mathbf{A}|$, $\mathrm{cov}(\mathbf{A})$, $\mathrm{Tr}(\mathbf{A})$ and $(\mathbf{A})^H$ indicate the determinant, covariance, trace and Hermitian of $\mathbf{A}$. $\mathrm{diag}(\mathbf{A})$ represents a diagonal matrix containing the elements along the diagonal of $\mathbf{A}$. $[\mathbf{A}]_m$ refers to the mth element along the diagonal of $\mathbf{A}$. $\mathrm{vec}(\mathbf{A})$ is a vector obtained by stacking the columns of $\mathbf{A}$. $\lfloor \mathbf{A}_k \rfloor_{k=1\ldots K}$ denotes a matrix obtained by stacking $\mathbf{A}_1, \ldots, \mathbf{A}_K$. $\otimes$, $\|\cdot\|$ and $\|\cdot\|_F$ indicate the Kronecker product, the Euclidean norm and the Frobenius norm.

II. Preliminaries

A. System model

We consider a scenario having G cells, where each cell g has one FD BS, K_g^d DL users requiring b_d streams each and K_g^u UL users requiring b_u streams each. BSs are equipped with M_B FD antennas, DL users are equipped with M_d HD antennas and UL users are equipped with M_u HD antennas. The maximum transmit power is given by P_B at each BS and P_U at each of the UL users.

The signal received at user k_g^d, the kth DL user in cell g, and at BS g are given by (1) and (2) respectively. Here, $\mathbf{H}_{k_g^d,j} \in \mathbb{C}^{M_d \times M_B}$ represents the channel from BS j to DL user k_g^d, $\mathbf{H}_{k_g^d,i_j^u} \in \mathbb{C}^{M_d \times M_u}$ is the channel from UL user i_j^u to DL user k_g^d, $\mathbf{H}_{g,j} \in \mathbb{C}^{M_B \times M_B}$ is the channel from BS j to BS g and $\mathbf{H}_{g,i_j^u} \in \mathbb{C}^{M_B \times M_u}$ is the channel from UL user i_j^u to BS g. $\mathbf{V}_{i_j^d} \in \mathbb{C}^{M_B \times b_d}$ is the precoder for $\mathbf{s}_{i_j^d}$, with $\mathbf{s}_{i_j^d} \in \mathbb{C}^{b_d \times 1}$ being the data intended for the ith DL user in cell j, where $\mathbb{E}[\mathbf{s}_{i_j^d}\mathbf{s}_{i_j^d}^H] = \mathbf{I}$. $\mathbf{V}_{i_j^u} \in \mathbb{C}^{M_u \times b_u}$ is the precoder for $\mathbf{s}_{i_j^u}$, with $\mathbf{s}_{i_j^u} \in \mathbb{C}^{b_u \times 1}$ being the data transmitted by the ith UL user in cell j, where $\mathbb{E}[\mathbf{s}_{i_j^u}\mathbf{s}_{i_j^u}^H] = \mathbf{I}$. Moreover, $\mathbf{n}_{k_g^d}$ and $\mathbf{n}_g$ represent additive white Gaussian noise with zero mean and variance σ_U^2 and σ_B^2 respectively. Finally, $\mathbf{c}_{i_j^u}$ and $\mathbf{c}_{i_j^d}$ represent transmitter distortion at UL users and at the BSs respectively, while $\mathbf{e}_{k_g^d}$ and $\mathbf{e}_g$ represent receiver distortion at DL users and at the BSs respectively.

Transmitter distortion models the effect of limited transmitter dynamic range by approximating the combined effects of additive power-amplifier noise, oscillator phase noise, and non-linearities in the DAC and the power amplifier. This distortion is statistically independent from the transmitted signal and can be modeled as [7]

$$\mathbf{c}_{i_j^d} \sim \mathcal{CN}\big(\mathbf{0}, \kappa_B \,\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\big)$$
$$\mathbf{c}_{i_j^u} \sim \mathcal{CN}\big(\mathbf{0}, \kappa_U \mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\big)$$

$$\mathbf{y}_{k_g^d} = \sum_{j=1}^{G} \mathbf{H}_{k_g^d,j} \sum_{i=1}^{K_j^d} (\mathbf{V}_{i_j^d}\mathbf{s}_{i_j^d} + \mathbf{c}_{i_j^d}) + \sum_{j=1}^{G} \sum_{i=1}^{K_j^u} \mathbf{H}_{k_g^d,i_j^u} (\mathbf{V}_{i_j^u}\mathbf{s}_{i_j^u} + \mathbf{c}_{i_j^u}) + \mathbf{n}_{k_g^d} + \mathbf{e}_{k_g^d} \tag{1}$$

$$\mathbf{y}_{g} = \sum_{j=1}^{G} \mathbf{H}_{g,j} \sum_{i=1}^{K_j^d} (\mathbf{V}_{i_j^d}\mathbf{s}_{i_j^d} + \mathbf{c}_{i_j^d}) + \sum_{j=1}^{G} \sum_{i=1}^{K_j^u} \mathbf{H}_{g,i_j^u} (\mathbf{V}_{i_j^u}\mathbf{s}_{i_j^u} + \mathbf{c}_{i_j^u}) + \mathbf{n}_{g} + \mathbf{e}_{g} \tag{2}$$

$$\tilde{\mathbf{y}}_{g} = \sum_{\substack{j=1 \\ j\neq g}}^{G} \mathbf{H}_{g,j} \sum_{i=1}^{K_j^d} (\mathbf{V}_{i_j^d}\mathbf{s}_{i_j^d} + \mathbf{c}_{i_j^d}) + \sum_{j=1}^{G} \sum_{i=1}^{K_j^u} \mathbf{H}_{g,i_j^u} (\mathbf{V}_{i_j^u}\mathbf{s}_{i_j^u} + \mathbf{c}_{i_j^u}) + \mathbf{n}_{g} + \mathbf{e}_{g} + \mathbf{H}_{g,g} \sum_{i=1}^{K_g^d} \mathbf{c}_{i_j^d} + \underbrace{\Theta \sum_{i=1}^{K_g^d} \mathbf{\Delta}_{g,g} \mathbf{V}_{i_j^d}\mathbf{s}_{i_j^d}}_{\text{extra residual SI for imperfect CSI scenarios}} \tag{3}$$

where $\kappa_U, \kappa_B \ll 1$.

Receiver distortion models the effect of limited receiver dynamic range by capturing the combined effects of oscillator phase noise, additive gain control noise, and non-linearities in the ADC and gain-control. It is statistically independent from the received signal itself and can be modeled as [7]

$$\mathbf{e}_{k_g^d} \sim \mathcal{CN}\Big(\mathbf{0}, \beta_U \mathrm{diag}\big(\mathrm{cov}(\mathbf{y}_{k_g^d} - \mathbf{e}_{k_g^d})\big)\Big)$$
$$\mathbf{e}_{g} \sim \mathcal{CN}\Big(\mathbf{0}, \beta_B \mathrm{diag}\big(\mathrm{cov}(\mathbf{y}_{g} - \mathbf{e}_{g})\big)\Big)$$

where $\beta_U, \beta_B \ll 1$.

Finally, since with perfect CSI $\mathbf{H}_{g,g} \sum_{i=1}^{K_g^d} \mathbf{V}_{i_g^d}\mathbf{s}_{i_g^d}$ is known at BS g, this can be subtracted from $\mathbf{y}_g$ resulting in (3) with $\Theta = 0$ [7][1,2]. The parameter Θ is a binary term used to differentiate between the perfect and imperfect CSI scenarios. For the perfect CSI case $\Theta = 0$, whilst for the imperfect CSI case $\Theta = 1$ leading to an extra residual SI term; further details for the imperfect CSI case are provided in Section II-B.

Similar to prior work dealing with beamforming and interference management, our proposed algorithms require CSI knowledge in order to be implemented. While going into the exact details is beyond the scope of this work, it is important to highlight the fact that all relevant channels can indeed be learned. Channels from users to BSs, from BSs to users and between BSs can be estimated using standard 3GPP LTE channel estimation protocols for HD systems. Channels between the users can be learned via neighbour discovery methods applicable to device-to-device (D2D) communication, such as sounding reference signals (SRS) in 3GPP LTE. (See for example [12], [24], [28] and references therein for further details on channel estimation.)

B. *Imperfect CSI considerations*

Whilst perfect CSI formulations provide a useful baseline to highlight the advantages of FD over HD, it is important to recognise that the perfect CSI assumption is idealistic; in practice only an imperfect estimate will be available. Therefore moving on from the original perfect CSI assumption we will also consider the design of robust beamformers. The channels are modeled as

$$\begin{aligned}
\mathbf{H}_{k_g^d,i_j^u} &= \hat{\mathbf{H}}_{k_g^d,i_j^u} + \Delta_{k_g^d,i_j^u} \\
\mathbf{H}_{k_g^d,j} &= \hat{\mathbf{H}}_{k_g^d,j} + \Delta_{k_g^d,j} \\
\mathbf{H}_{g,i_j^u} &= \hat{\mathbf{H}}_{g,i_j^u} + \Delta_{g,i_j^u} \\
\mathbf{H}_{g,j} &= \hat{\mathbf{H}}_{g,j} + \Delta_{g,j}
\end{aligned} \tag{4}$$

where $\mathbf{H}$ indicates the perfect channel, $\hat{\mathbf{H}}$ is the imperfect channel and Δ is the CSI error. Note that the significance of the indices used to represent the channels and errors between various nodes in (4) follow those outlined for (1) and (2).

For the imperfect CSI case only $\hat{\mathbf{H}}_{g,g} \sum_{i=1}^{K_g^d} \mathbf{V}_{i_g^d}\mathbf{s}_{i_g^d}$ is known at BS g. This can be subtracted from $\mathbf{y}_g$ resulting in (3) with $\Theta = 1$, where there is an extra residual SI component compared to the perfect CSI case.

The CSI error will be modeled in two different ways as follows.

1) Norm-bounded error model: For the deterministic norm-bounded error model, the Frobenius norm of the CSI errors cannot exceed a pre-established upper bound, and the CSI error is expressed as

$$\begin{aligned}
&\{\Delta_{k_g^d,i_j^u} : ||\Delta_{k_g^d,i_j^u}||_F \leq \varepsilon_{k_g^d,i_j^u}\} \;\; \forall\, k,g,i,j \\
&\{\Delta_{k_g^d,j} : ||\Delta_{k_g^d,j}||_F \leq \varepsilon_{k_g^d,j}\} \;\; \forall\, k,g,j \\
&\{\Delta_{g,i_j^u} : ||\Delta_{g,i_j^u}||_F \leq \varepsilon_{g,i_j^u}\} \;\; \forall\, g,i,j \\
&\{\Delta_{g,j} : ||\Delta_{g,j}||_F \leq \varepsilon_{g,j}\} \;\; \forall\, g,j
\end{aligned} \tag{5}$$

where ε represents the upper limit on the Frobenius norm of the error. This model considers the case where the imperfect CSI is allowed to fall anywhere within an uncertainty region around the perfect CSI value and is particularly suited to situations where quantization errors dominate the imperfection in the available CSI. It is well established in literature and has been considered for beamformer design in a variety of systems, for example MIMO relay networks [29], MIMO interference broadcast channels [30], DL multi-user MIMO systems [31] and point-to-point MIMO communication [32], [33].

2) Stochastic error model: For the stochastic error model the CSI errors are assumed to be independent of the perfect channel, $\mathbf{H}$, and distributed as follows

$$\begin{aligned}
\Delta_{k_g^d,i_j^u} &\sim \mathcal{CN}(0, \eta_{UU}\mathbf{I}) \\
\Delta_{k_g^d,j} &\sim \mathcal{CN}(0, \eta_{UB}\mathbf{I}) \\
\Delta_{g,i_j^u} &\sim \mathcal{CN}(0, \eta_{BU}\mathbf{I})
\end{aligned}$$

[1] The SI channel can be estimated using pilot signals. For SI channel estimation, the FD node transmitting the pilot signal is also the one receiving it, this implies that the signal is received with high power. Having a strong signal allows for accurate estimation of the SI channel $\mathbf{H}_{g,g}$ [24], which implies that $\mathbf{H}_{g,g} \sum_{i=1}^{K_g^d} \mathbf{V}_{i_g^d}\mathbf{s}_{i_g^d}$ can be considered as available at BS g under the perfect CSI assumption. The effect of residual SI is then captured in the term $\mathbf{H}_{g,g} \sum_{i=1}^{K_g^d} \mathbf{c}_{i_j^d} + \mathbf{e}_g$.

[2] Note that other literature like [7], [8], [20], [25] adopts a partial SI cancellation method similar to the one used in this paper, while others implement partial SI cancellation via the use of an attenuation factor [26]. However a completely different model may also be adopted where $H_{g,g}$ is used to represent the residual SI channel directly, see for example [27].

$$\mathbf{F}_{k_g^d} \approx \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \kappa_B \mathbf{H}_{k_g^d,j}\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\mathbf{H}_{k_g^d,j}^H + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \kappa_U \mathbf{H}_{k_g^d,i_j^u}\mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\mathbf{H}_{k_g^d,i_j^u}^H + \beta_U\sigma_U^2\mathbf{I}$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \beta_U \mathrm{diag}(\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{k_g^d,j}^H) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \beta_U \mathrm{diag}(\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H) \quad (7)$$

$$\bar{\mathbf{U}}_{k_g^d} = \mathbf{V}_{k_g^d}^H\mathbf{H}_{k_g^d,g}^H\left[\sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{k_g^d,j}^H + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H + \mathbf{F}_{k_g^d} + \sigma_U^2\mathbf{I}\right]^{-1} \quad (8)$$

$$\mathbf{F}_g \approx \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \kappa_B \mathbf{H}_{g,j}\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\mathbf{H}_{g,j}^H + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \kappa_U \mathbf{H}_{g,i_j^u}\mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\mathbf{H}_{g,i_j^u}^H + \beta_B\sigma_B^2\mathbf{I}$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \beta_B \mathrm{diag}(\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \beta_B \mathrm{diag}(\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{g,i_j^u}^H) \quad (10)$$

$$\bar{\mathbf{U}}_{k_g^u} = \mathbf{V}_{k_g^u}^H\mathbf{H}_{g,k_g^u}^H\left[\sum_{\substack{j=1\\ j\neq g}}^{G}\sum_{i=1}^{K_j^d}\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{g,i_j^u}^H + \mathbf{F}_g + \sigma_B^2\mathbf{I}\right]^{-1} \quad (11)$$

$$\Delta_{g,j} \sim \mathcal{CN}(0, \eta_{BB}\mathbf{I}) \quad (6)$$

where η represents the variance of the CSI error and the subscripts B and U indicate the BS and user respectively. This type of error model is suitable for cases where the channel error is mainly due to estimation inaccuracies. The parameter η can be assumed to be known *a priori* depending on the channel dynamics and the channel estimation scheme applied. It may be viewed either as a whole [34] or modeled as [35]

$$\eta_{rt} = \tau\rho_{rt}^{-\nu}$$

where $r,t \in \{B,U\}$, ρ represents the signal-to-noise (SNR) ratio of the corresponding link and the parameters τ and ν are used to capture a variety of CSI acquisition scenarios for $\tau > 0$ and $\nu \geq 0$.

C. Relationship between achievable rate and MSE

Since our approach for WSR maximization is based on minimizing the mean square error (MSE), we first need to relate these two metrics. Such a relationship has already been established for several HD systems in [21]–[23], where it is shown that $R = \log_2|\bar{\mathbf{E}}^{-1}|$, with R representing the rate and $\bar{\mathbf{E}}$ representing the MSE matrix[3]. This equality holds for independent input signals and noise, and for cases where optimal MMSE receivers are used.

Starting with the rate expression for the kth DL user in cell g, under Gaussian signaling, we have $R_{k_g^d} = \log_2|\mathbf{I} + \mathbf{\Phi}_{k_g^d}^{-1}\mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H\mathbf{H}_{k_g^d,g}^H|$ where $\mathbf{\Phi}_{k_g^d}$ represents the DL interference-plus-noise covariance matrix, expressed as

$$\mathbf{\Phi}_{k_g^d} = \sum_{\substack{j=1\\(i,j\neq k,g)}}^{G}\sum_{i=1}^{K_j^d}\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{k_g^d,j}^H + \mathbf{F}_{k_g^d} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H + \sigma_U^2\mathbf{I}$$

with $\mathbf{F}_{k_g^d}$, defined in (7), representing the combined contribution of the transmitter and receiver distortion. The approximation is obtained by omitting terms involving the multiplication of κ_B, κ_U and β_U with each other since their product is negligibly small.

The DL MSE matrix is given by $\mathbf{E}_{k_g^d} = \mathbb{E}[(\mathbf{U}_{k_g^d}\mathbf{y}_{k_g^d} - \mathbf{s}_{k_g^d})(\mathbf{U}_{k_g^d}\mathbf{y}_{k_g^d} - \mathbf{s}_{k_g^d})^H]$ where the expectation is taken with respect to $\mathbf{s}$ and $\mathbf{n}$ under an independence assumption, and $\mathbf{U}_{k_g^d} \in \mathbb{C}^{b_d \times M_d}$ is the receiver applied by the kth DL user in cell g. Applying an MMSE receiver, $\bar{\mathbf{U}}_{k_g^d} = \arg\min_{\mathbf{U}} \mathrm{Tr}(\mathbf{E}_{k_g^d})$ given by (8), then the DL MSE matrix can be expressed as $\bar{\mathbf{E}}_{k_g^d} = (\mathbf{I} + \mathbf{V}_{k_g^d}^H\mathbf{H}_{k_g^d,g}^H\mathbf{\Phi}_{k_g^d}^{-1}\mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d})^{-1}$. Finally using an argument parallel to the one from [21]–[23], it can be established that

$$R_{k_g^d} = \log_2\left|\bar{\mathbf{E}}_{k_g^d}^{-1}\right| . \quad (9)$$

The rate achieved by the kth UL user in cell g, under Gaussian signaling, can be expressed as $R_{k_g^u} = \log_2|\mathbf{I} + \mathbf{\Phi}_{k_g^u}^{-1}\mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H\mathbf{H}_{g,k_g^u}^H|$ where $\mathbf{\Phi}_{k_g^u}$ represents the UL interference-plus-noise covariance matrix, given by

$$\mathbf{\Phi}_{k_g^u} = \sum_{\substack{j=1\\(j\neq g)}}^{G}\sum_{i=1}^{K_j^d}\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H + \mathbf{F}_g + \sum_{\substack{j=1\\(i,j\neq k,g)}}^{G}\sum_{i=1}^{K_j^u}\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{g,i_j^u}^H + \sigma_B^2\mathbf{I}$$

with $\mathbf{F}_g$, defined in (10), representing the effect of transmitter and receiver distortion.

The UL MSE matrix is given by $\mathbf{E}_{k_g^u} = \mathbb{E}[(\mathbf{U}_{k_g^u}\tilde{\mathbf{y}}_{k_g^u} - \mathbf{s}_{k_g^u})(\mathbf{U}_{k_g^u}\tilde{\mathbf{y}}_{k_g^u} - \mathbf{s}_{k_g^u})^H]$ where the expectation is taken with respect to $\mathbf{s}$ and $\mathbf{n}$ under an independence assumption, and $\mathbf{U}_{k_g^u} \in \mathbb{C}^{b_u \times M_B}$ is the receiver applied by BS g to obtain the information transmitted by the kth UL in its cell. Applying an MMSE receiver, $\bar{\mathbf{U}}_{k_g^u} = \arg\min_{\mathbf{U}} \mathrm{Tr}(\mathbf{E}_{k_g^u})$ given by (11), then the UL MSE matrix can then be expressed as $\bar{\mathbf{E}}_{k_g^u} = \mathbf{I} + \mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u}\mathbf{\Phi}_{k_g^u}^{-1}\mathbf{V}_{k_g^u}^H\mathbf{H}_{g,k_g^u}^H$. Next using an argument similar to the one from [21]–[23], we can establish that

$$R_{k_g^u} = \log_2\left|\bar{\mathbf{E}}_{k_g^u}^{-1}\right| . \quad (12)$$

[3]Note that [21] and [22] establish the rate to MSE relationship using base e logarithms, while this paper and [23] establish it using base 2 logarithms. The overall rate to MSE relationship is essentially the same, with the choice of logarithm only affecting the corresponding unit for R. For base e logarithms R is in nats/s/Hz, while for base 2 logarithms R is in bits/s/Hz. The resultant WSR to WMMSE problem transformations are also similar, except for the ln2 terms which do not appear when base e logarithms are applied.

$$\mathrm{Tr}\left(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}\right) = \mathrm{Tr}\left(\mathbf{B}_{k_g^d}(\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d} - \mathbf{I})(\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,g}\mathbf{V}_{k_g^d} - \mathbf{I})^H\mathbf{B}_{k_g^d}^H\right) + (\sigma_U^2 + \beta_U\sigma_U^2)\mathrm{Tr}\left(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathrm{Tr}\left(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right) + \sum_{\substack{j=1\\(i,j\neq k,g)}}^{G}\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{k_g^d,j}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\kappa_B\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j}\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\,\mathbf{H}_{k_g^d,j}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathrm{Tr}\left(\kappa_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u}\,\mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\beta_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathrm{diag}(\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{k_g^d,j}^H)\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathrm{Tr}\left(\beta_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathrm{diag}(\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H)\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right) \quad (15)$$

$$\mathrm{Tr}\left(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}\right) = \mathrm{Tr}\left(\mathbf{B}_{k_g^u}(\mathbf{U}_{k_g^u}\mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u} - \mathbf{I})(\mathbf{U}_{k_g^u}\mathbf{H}_{g,k_g^u}\mathbf{V}_{k_g^u} - \mathbf{I})\mathbf{B}_{k_g^u}^H\right) + (\sigma_B^2 + \beta_B\sigma_B^2)\mathrm{Tr}\left(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)$$
$$+ \sum_{\substack{j=1\\(i,j\neq k,g)}}^{G}\sum_{i=1}^{K_j^u}\mathrm{Tr}\left(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{g,i_j^u}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right) + \sum_{\substack{j=1\\(j\neq g)}}^{G}\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\kappa_B\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,j}\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\mathbf{H}_{g,j}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathrm{Tr}\left(\kappa_U\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u}\mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\mathbf{H}_{g,i_j^u}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)$$
$$+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\beta_B\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathrm{diag}(\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H)\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u}\mathrm{Tr}\left(\beta_B\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathrm{diag}(\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{g,i_j^u}^H)\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)$$
$$+ \Theta\sum_{i=1}^{K_j^d}\mathrm{Tr}\left(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\Delta_{g,g}\mathbf{V}_{i_g^d}\mathbf{V}_{i_g^d}^H\Delta_{g,g}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right) \quad (16)$$

III. Weighted sum rate maximization

Starting with the perfect CSI case, we want to find the optimal precoders that maximize the WSR subject to transmit power constraints, i.e.

$$\begin{aligned}\max_{\mathbf{V}} &\sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d}\alpha_{k_g^d}R_{k_g^d} + \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^u}\alpha_{k_g^u}R_{k_g^u}\\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \quad \forall\, k,g\\ & \sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \quad \forall\, g\end{aligned} \quad (13)$$

where $\alpha_{k_g^d}$ and $\alpha_{k_g^u}$ $\forall\, k,g$ denote pre-defined weights. Using a method parallel to that of [21]–[23] we can establish an equivalence between the WSR problem (13) and a corresponding WMMSE one as in Theorem 1.

Theorem 1. *The WSR problem in (13) is equivalent to the WMMSE problem in (14), such that the global optimal solution for the precoders of the two problems are identical.*

$$\begin{aligned}\min_{\mathbf{U},\mathbf{W},\mathbf{V}} &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\left[\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) - \alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{W}_{k_g^d}\right| - \frac{\alpha_{k_g^d}}{\ln 2}b_d\right]\\ &+\sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left[\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) - \alpha_{k_g^u}\log_2\left|\frac{\ln 2}{\alpha_{k_g^u}}\mathbf{W}_{k_g^u}\right| - \frac{\alpha_{k_g^u}}{\ln 2}b_u\right]\\ \textit{s.t.} \quad & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \quad \forall\, k,g\\ & \sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \quad \forall\, g\end{aligned} \quad (14)$$

Proof. First we define the metrics $\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d})$ and $\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u})$ used in (14) as in (15) and (16). Here $\mathbf{B}_{k_g^d}$ comes from the decomposition of $\mathbf{W}_{k_g^d}$ as $\mathbf{W}_{k_g^d} = \mathbf{B}_{k_g^d}^H\mathbf{B}_{k_g^d}$ and $\mathbf{B}_{k_g^u}$ comes from the decomposition of $\mathbf{W}_{k_g^u}$ as $\mathbf{W}_{k_g^u} = \mathbf{B}_{k_g^u}^H\mathbf{B}_{k_g^u}$.

Considering (14), it can be seen that the optimal $\mathbf{U}$ are the standard MMSE receivers $\bar{\mathbf{U}}_{k_g^d}$ and $\bar{\mathbf{U}}_{k_g^u}$ in (8) and (11) respectively. Next, fixing $\mathbf{U}$ and $\mathbf{V}$ and checking the first order optimality conditions for $\mathbf{W}$, we obtain the optimal weights as

$$\bar{\mathbf{W}}_{k_g^d} = \frac{\alpha_{k_g^d}}{\ln 2}\bar{\mathbf{E}}_{k_g^d}^{-1} \text{ and } \bar{\mathbf{W}}_{k_g^u} = \frac{\alpha_{k_g^u}}{\ln 2}\bar{\mathbf{E}}_{k_g^u}^{-1}\ . \quad (17)$$

Substituting for optimal $\mathbf{U}$ and $\mathbf{W}$ in (14), we have the following problem

$$\begin{aligned}\min_{\mathbf{V}} &\sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d}-\alpha_{k_g^d}\log_2\left|\bar{\mathbf{E}}_{k_g^d}^{-1}\right| + \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^u}-\alpha_{k_g^u}\log_2\left|\bar{\mathbf{E}}_{k_g^u}^{-1}\right|\\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \quad \forall\, k,g\\ & \sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \quad \forall\, g\end{aligned}$$

which considering (9) and (12) is the same as the original one in (13). □

Since (14) is not jointly convex in $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ but separately convex for each of the variables, we apply an alternating minimization approach to solve the problem as outlined in Algorithm 1. Having closed form expressions for $\mathbf{U}$ and $\mathbf{W}$, we need to focus on obtaining $\mathbf{V}$. Fixing $\mathbf{U}$ and $\mathbf{W}$, (14) can be expressed as

$$\begin{aligned}\min_{\mathbf{V}} &\sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) + \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^u}\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u})\\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \quad \forall\, k,g\end{aligned}$$

$$\mathcal{L} = \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) + \sum_{g=1}^{G}\sum_{k=1}^{K_g^u} \lambda_{k_g^u}\left[\mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) - P_U\right] + \sum\nolimits_{g=1}^{G} \mu_g \left[\sum\nolimits_{k=1}^{K_g^d} \mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) - P_B\right] \tag{19}$$

$$\begin{aligned}\mathbf{X}_g = &\sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \mathbf{H}_{i_j^d,g}^H \mathbf{U}_{i_j^d}^H \mathbf{W}_{i_j^d}\mathbf{U}_{i_j^d}\mathbf{H}_{i_j^d,g} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \kappa_{BS}\mathrm{diag}(\mathbf{H}_{i_j^d,g}^H \mathbf{U}_{i_j^d}^H \mathbf{W}_{i_j^d}\mathbf{U}_{i_j^d}\mathbf{H}_{i_j^d,g}) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \beta_{us}\mathbf{H}_{i_j^d,g}^H \mathrm{diag}(\mathbf{U}_{i_j^d}^H \mathbf{W}_{i_j^d}\mathbf{U}_{i_j^d})\mathbf{H}_{i_j^d,g}\\ &+ \sum_{\substack{j=1\\ j\neq g}}^{G}\sum_{i=1}^{K_j^u} \mathbf{H}_{j,g}^H \mathbf{U}_{i_j^u}^H \mathbf{W}_{i_j^u}\mathbf{U}_{i_j^u}\mathbf{H}_{j,g} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \kappa_{BS}\mathrm{diag}(\mathbf{H}_{j,g}^H \mathbf{U}_{i_j^u}^H \mathbf{W}_{i_j^u}\mathbf{U}_{i_j^u}\mathbf{H}_{j,g}) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \beta_{BS}\mathbf{H}_{j,g}^H diag(\mathbf{U}_{i_j^u}^H \mathbf{W}_{i_j^u}\mathbf{U}_{i_j^u})\mathbf{H}_{j,g}\end{aligned} \tag{21}$$

$$\begin{aligned}\mathbf{X}_{k_g^u} = &\sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \mathbf{H}_{i_j^d,k_g^u}^H \mathbf{U}_{i_j^d}^H \mathbf{W}_{i_j^d}\mathbf{U}_{i_j^d}\mathbf{H}_{i_j^d,k_g^u} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \kappa_{us}\mathrm{diag}(\mathbf{H}_{i_j^d,k_g^u}^H \mathbf{U}_{i_j^d}^H \mathbf{W}_{i_j^d}\mathbf{U}_{i_j^d}\mathbf{H}_{i_j^d,k_g^u}) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \beta_{us}\mathbf{H}_{i_j^d,k_g^u}^H \mathrm{diag}(\mathbf{U}_{i_j^d}^H \mathbf{W}_{i_j^d}\mathbf{U}_{i_j^d})\mathbf{H}_{i_j^d,k_g^u}\\ &+ \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \mathbf{H}_{j,k_g^u}^H \mathbf{U}_{i_j^u}^H \mathbf{W}_{i_j^u}\mathbf{U}_{i_j^u}\mathbf{H}_{j,k_g^u} + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \kappa_{us}\mathrm{diag}(\mathbf{H}_{j,k_g^u}^H \mathbf{U}_{i_j^u}^H \mathbf{W}_{i_j^u}\mathbf{U}_{i_j^u}\mathbf{H}_{j,k_g^u}) + \sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \beta_{BS}\mathbf{H}_{j,k_g^u}^H \mathrm{diag}(\mathbf{U}_{i_j^u}^H \mathbf{W}_{i_j^u}\mathbf{U}_{i_j^u})\mathbf{H}_{j,k_g^u}\end{aligned} \tag{22}$$

$$\sum\nolimits_{k=1}^{K_g^d} \mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \ \ \forall\, g\ . \tag{18}$$

The Lagrange dual objective function of (18) is given by (19) where $\lambda_{k_g^u}$ and μ_g are the Lagrange multipliers associated with the transmit power constraints. Setting $\partial\mathcal{L}/\partial\mathbf{V}_{k_g^d}^* = 0$ and $\partial\mathcal{L}/\partial\mathbf{V}_{k_g^u}^* = 0$ we obtain the closed form solutions for the optimal precoders as

$$\begin{aligned}\bar{\mathbf{V}}_{k_g^d} &= [\mathbf{X}_g + \mu_g\mathbf{I}]^{-1}\mathbf{H}_{k_g^d,g}^H\mathbf{U}_{k_g^d}^H\mathbf{W}_{k_g^d}\\ \bar{\mathbf{V}}_{k_g^u} &= [\mathbf{X}_{k_g^u} + \lambda_{k_g^u}\mathbf{I}]^{-1}\mathbf{H}_{g,k_g^u}^H\mathbf{U}_{k_g^u}^H\mathbf{W}_{k_g^u}\end{aligned} \tag{20}$$

where $\mathbf{X}_g$ and $\mathbf{X}_{k_g^u}$ are defined in (21) and (22). The Lagrange multipliers μ_g and $\lambda_{k_g^u}$ should be either zero, or positive numbers that satisfy

$$\begin{aligned}\mathrm{Tr}\left(\mathbf{V}_{k_g^u}(\lambda_{k_g^u})\mathbf{V}_{k_g^u}^H(\lambda_{k_g^u})\right) &= P_U\\ \sum\nolimits_{k=1}^{K_g^d} \mathrm{Tr}\left(\mathbf{V}_{k_g^d}(\mu_g)\mathbf{V}_{k_g^d}^H(\mu_g)\right) &= P_B\ .\end{aligned} \tag{23}$$

The equalities in (23) can alternatively be expressed as

$$\begin{aligned}\sum\nolimits_{m=1}^{M_u} \frac{[\mathbf{G}_{k_g^u}]_m}{([\mathbf{\Lambda}_{k_g^u}]_m + \lambda_{k_g^u})^2} &= P_U\\ \sum\nolimits_{m=1}^{M_B} \frac{[\mathbf{G}_g]_m}{([\mathbf{D}_g]_m + \mu_g)^2} &= P_B\end{aligned}$$

where $\mathbf{\Lambda}_{k_g^u}$ comes from the decomposition $\mathbf{X}_{k_g^u} = \mathbf{C}_{k_g^u}\mathbf{\Lambda}_{k_g^u}\mathbf{C}_{k_g^u}^H$, $\mathbf{G}_{k_g^u} = \mathbf{C}_{k_g^u}^H\mathbf{H}_{g,k_g^u}^H\mathbf{U}_{k_g^u}^H\mathbf{W}_{k_g^u}\mathbf{W}_{k_g^u}^H\mathbf{U}_{k_g^u}\mathbf{H}_{g,k_g^u}\mathbf{C}_{k_g^u}$, $\mathbf{D}_g$ comes from the decomposition $\mathbf{X}_g = \mathbf{Q}_g\mathbf{D}_g\mathbf{Q}_g^H$ and $\mathbf{G}_g = \sum_{k=1}^{K_g^d}\mathbf{Q}_g^H\mathbf{H}_{k_g^d,g}^H\mathbf{U}_{k_g^d}^H\mathbf{W}_{k_g^d}\mathbf{W}_{k_g^d}^H\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,g}\mathbf{Q}_g$. These can respectively be solved for $\lambda_{k_g^u}$ and μ_g using linear search techniques, such as the bisection method [22].

Therefore to solve (14), we can follow the process in Algorithm 1, where in Step 2 we use (8) and (11) to calculate the receivers as $\mathbf{U}_{k_g^d} = \bar{\mathbf{U}}_{k_g^d}$ and $\mathbf{U}_{k_g^u} = \bar{\mathbf{U}}_{k_g^u}$. The weights in Step 3 are calculated as $\mathbf{W}_{k_g^d} = \bar{\mathbf{W}}_{k_g^d}$ and $\mathbf{W}_{k_g^u} = \bar{\mathbf{W}}_{k_g^u}$ using (17). Finally (20) is used to calculate $\mathbf{V}_{k_g^d} = \bar{\mathbf{V}}_{k_g^d}$ and $\mathbf{V}_{k_g^u} = \bar{\mathbf{V}}_{k_g^u}$ in Step 4.

Remark 1. *The alternating minimization process used to solve the WMMSE problem decreases the cost function monotonically at each step of the iterations. Since the cost function is lower bounded, then the algorithm is guaranteed to converge. Additionally, using an argument parallel to the one in [22]-Appendix C, convergence to a stationary point of the original WSR problem can also be proven.*

Algorithm 1: Alternating optimization process to solve WMMSE problems

1 Initialize $\mathbf{V}_{k_g^d}$ and $\mathbf{V}_{k_g^u}$ $\forall\, k, g$.
2 Calculate $\mathbf{U}_{k_g^d}$ and $\mathbf{U}_{k_g^u}$ $\forall\, k, g$.
3 Calculate $\mathbf{W}_{k_g^d}$ and $\mathbf{W}_{k_g^u}$ $\forall\, k, g$.
4 Compute $\mathbf{V}_{k_g^d}$ and $\mathbf{V}_{k_g^u}$ $\forall\, k, g$.
5 Repeat from Step 2 until convergence or for a fixed number of iterates.

IV. Robust design with norm-bounded error model

We now want to solve the WSR problem from the prior section with additional considerations for norm-bounded CSI errors, i.e.

$$\begin{aligned}\max_{\mathbf{V}}\min_{\Delta}\ & \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d}\alpha_{k_g^d}R_{k_g^d} + \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^u}\alpha_{k_g^u}R_{k_g^u}\\ \text{s.t.}\ & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \ \ \forall\, k, g\\ & \sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \ \ \forall\, g\\ & \{\Delta_{k_g^d,i_j^u} : ||\Delta_{k_g^d,i_j^u}||_F \leq \varepsilon_{k_g^d,i_j^u}\} \ \ \forall\, k, g, i, j\\ & \{\Delta_{k_g^d,j} : ||\Delta_{k_g^d,j}||_F \leq \varepsilon_{k_g^d,j}\} \ \ \forall\, k, g, j\\ & \{\Delta_{g,i_j^u} : ||\Delta_{g,i_j^u}||_F \leq \varepsilon_{g,i_j^u}\} \ \ \forall\, g, i, j\\ & \{\Delta_{g,j} : ||\Delta_{g,j}||_F \leq \varepsilon_{g,j}\} \ \ \forall\, g, j\ .\end{aligned} \tag{24}$$

Similar to [30], [31] and references therein, we apply an iterative approach to solve our non-convex optimization problem. Such an approach involves solving a convex sub-problem at each iteration step and has been proven to converge. Having already established an equivalence between (13) and

(14) for the perfect CSI case, it can easily bee seen how the cost function of (24) can be mapped to

$$\max_{\mathbf{V}} \min_{\Delta} \max_{\mathbf{U},\mathbf{W}} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \left[-\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) + \alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{W}_{k_g^d}\right| + \frac{\alpha_{k_g^d}}{\ln 2}b_d\right]$$
$$+\sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left[-\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) + \alpha_{k_g^u}\log_2\left|\frac{\ln 2}{\alpha_{k_g^u}}\mathbf{W}_{k_g^u}\right| + \frac{\alpha_{k_g^u}}{\ln 2}b_u\right]. \tag{25}$$

Applying the max-min inequality, which states that for any function $f(w,z)$ then $\min_w \max_z f(w,z) \geq \max_z \min_w f(w,z)$, rather than using the cost function in (25) we can instead focus on solving the following problem

$$\max_{\mathbf{V},\mathbf{U},\mathbf{W}} \min_{\Delta} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} \left[-\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) + \alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{W}_{k_g^d}\right| + \frac{\alpha_{k_g^d}}{\ln 2}b_d\right]$$
$$+\sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left[-\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) + \alpha_{k_g^u}\log_2\left|\frac{\ln 2}{\alpha_{k_g^u}}\mathbf{W}_{k_g^u}\right| + \frac{\alpha_{k_g^u}}{\ln 2}b_u\right]$$
$$\text{s.t.} \quad \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \quad \forall\, k,g$$
$$\sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \quad \forall\, g$$
$$\{\Delta_{k_g^d,i_j^u} : ||\Delta_{k_g^d,i_j^u}||_F \leq \varepsilon_{k_g^d,i_j^u}\} \quad \forall\, k,g,i,j$$
$$\{\Delta_{k_g^d,j} : ||\Delta_{k_g^d,j}||_F \leq \varepsilon_{k_g^d,j}\} \quad \forall\, k,g,j$$
$$\{\Delta_{g,i_j^u} : ||\Delta_{g,i_j^u}||_F \leq \varepsilon_{g,i_j^u}\} \quad \forall\, g,i,j$$
$$\{\Delta_{g,j} : ||\Delta_{g,j}||_F \leq \varepsilon_{g,j}\} \quad \forall\, g,j\,. \tag{26}$$

Note that the cost function in (26) is not equivalent to the original one in (25). However the ensuing formulation is still a valid one; firstly the new cost function is a lower bound on (25) i.e. the resultant rate is surely achievable. Secondly the formulation in (26) ensures that none of the optimization variables depend on perfect CSI, which is the ultimate aim of a robust beamforming approach.

Theorem 2. *The optimization problem in (26) is equivalent to the reformulation in (27) such that the optimal* $\mathbf{U}$*,* $\mathbf{V}$ *and* $\mathbf{W} = \mathbf{B}^H\mathbf{B}$ *for the two problems are identical.*

$$\max_{\mathbf{V},\mathbf{U},\mathbf{B},m,\lambda} \sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\Bigg[-\sum_{j=1}^{G}\sum_{i=1}^{K_j^u} m_{1,k_g i_j} - \sum_{j=1}^{G} m_{2,k_g j}$$
$$-(\sigma_U^2+\beta_U\sigma_U^2)||\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}||_F^2 + \alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{B}_{k_g^d}^H\mathbf{B}_{k_g^d}\right| + \frac{\alpha_{k_g^d}}{\ln 2}b_d\Bigg]$$
$$+\sum_{g=1}^{G}\Bigg[-\sum_{j=1}^{G}\sum_{i=1}^{K_j^u} m_{3,g i_j} - \sum_{j=1}^{G} m_{4,gj} +$$
$$\sum_{k=1}^{K_g^u}\left(-(\sigma_B^2+\beta_B\sigma_B^2)||\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}||_F^2 + \alpha_{k_g^u}\log_2\left|\frac{\ln 2}{\alpha_{k_g^u}}\mathbf{B}_{k_g^u}^H\mathbf{B}_{k_g^u}\right| + \frac{\alpha_{k_g^u}}{\ln 2}b_u\right)\Bigg]$$
$$\textit{s.t.} \quad \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \leq P_U \;\forall\; k,g$$
$$\sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \leq P_B \;\forall\; g$$
$$\begin{bmatrix} m_{1,k_g i_j} - \lambda_{1,k_g i_j} & \boldsymbol{\omega}_{1,k_g i_j}^H & \mathbf{0} \\ \boldsymbol{\omega}_{1,k_g i_j} & \mathbf{I} & -\varepsilon_{k_g^d,i_j^u}\boldsymbol{\Omega}_{1,k_g i_j} \\ \mathbf{0} & -\varepsilon_{k_g^d,i_j^u}\boldsymbol{\Omega}_{1,k_g i_j}^H & \lambda_{1,k_g i_j}\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \;\forall\; k,g,i,j$$
$$\begin{bmatrix} m_{2,k_g j} - \lambda_{2,k_g j} & \boldsymbol{\omega}_{2,k_g j}^H & \mathbf{0} \\ \boldsymbol{\omega}_{2,k_g j} & \mathbf{I} & -\varepsilon_{k_g^d,j}\boldsymbol{\Omega}_{2,k_g j} \\ \mathbf{0} & -\varepsilon_{k_g^d,j}\boldsymbol{\Omega}_{2,k_g j}^H & \lambda_{2,k_g j}\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \;\forall\; k,g,j$$
$$\begin{bmatrix} m_{3,g i_j} - \lambda_{3,g i_j} & \boldsymbol{\omega}_{3,g i_j}^H & \mathbf{0} \\ \boldsymbol{\omega}_{3,g i_j} & \mathbf{I} & -\varepsilon_{g,i_j^u}\boldsymbol{\Omega}_{3,g i_j} \\ \mathbf{0} & -\varepsilon_{g,i_j^u}\boldsymbol{\Omega}_{3,g i_j}^H & \lambda_{3,g i_j}\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \;\forall\; g,i,j$$
$$\begin{bmatrix} m_{4,gj} - \lambda_{4,gj} & \boldsymbol{\omega}_{4,gj}^H & \mathbf{0} \\ \boldsymbol{\omega}_{4,gj} & \mathbf{I} & -\varepsilon_{g,j}\boldsymbol{\Omega}_{4,gj} \\ \mathbf{0} & -\varepsilon_{g,j}\boldsymbol{\Omega}_{4,gj}^H & \lambda_{4,gj}\mathbf{I} \end{bmatrix} \succeq \mathbf{0} \;\forall\; g,j$$
$$\lambda_{1,k_g i_j} \geq 0, \quad \lambda_{2,k_g j} \geq 0, \quad \lambda_{3,g i_j} \geq 0, \quad \lambda_{4,gj} \geq 0 \quad \forall\; k,g,i,j \tag{27}$$

In (27), m *and* λ *represent additional scalar variables introduced during the reformulation, and the* ω *and* Ω *terms are defined as follows.*

$$\boldsymbol{\omega}_{1k_g,i_j} = \begin{bmatrix} \mathrm{vec}(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\hat{\mathbf{H}}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}) \\ (\kappa_U)^{\frac{1}{2}}\left\lfloor\left((\mathbf{S}_n\mathbf{V}_{i_j^u})^T \otimes (\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d})\right)\mathrm{vec}(\hat{\mathbf{H}}_{k_g^d,i_j^u})\right\rfloor_{n=1\ldots M_u} \\ (\beta_U)^{\frac{1}{2}}\left\lfloor\left(\mathbf{V}_{i_j^u}^T \otimes \left((\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H)^T\mathbf{S}_n\right)\right)\mathrm{vec}(\hat{\mathbf{H}}_{k_g^d,i_j^u})\right\rfloor_{n=1\ldots M_d} \end{bmatrix}$$

$$\boldsymbol{\omega}_{2,k_g j} = \begin{bmatrix} \left\lfloor\mathrm{vec}(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\hat{\mathbf{H}}_{k_g^d,j}\mathbf{V}_{i_j^d} - \delta_{i,j}^{k,g}\mathbf{B}_{k_g^d})\right\rfloor_{i=1\ldots K_j^d} \\ \left\lfloor(\kappa_B)^{\frac{1}{2}}\left\lfloor\left((\mathbf{S}_n\mathbf{V}_{i_j^d})^T \otimes \mathrm{vec}(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d})\right)\mathrm{vec}(\hat{\mathbf{H}}_{k_g^d,j})\right\rfloor_{n=1\ldots M_B}\right\rfloor_{i=1\ldots K_j^d} \\ \left\lfloor(\beta_U)^{\frac{1}{2}}\left\lfloor\left(\mathbf{V}_{i_j^d}^T \otimes \left((\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H)^T\mathbf{S}_n\right)\right)\mathrm{vec}(\hat{\mathbf{H}}_{k_g^d,j})\right\rfloor_{n=1\ldots M_d}\right\rfloor_{i=1\ldots K_j^d} \end{bmatrix}$$

$$\boldsymbol{\omega}_{3,g i_j} = \begin{bmatrix} \left\lfloor\mathrm{vec}(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\hat{\mathbf{H}}_{g,i_j^u}\mathbf{V}_{i_j^u} - \delta_{i,j}^{k,g}\mathbf{B}_{k_g^d})\right\rfloor_{k=1\ldots K_g^u} \\ \left\lfloor(\kappa_U)^{\frac{1}{2}}\left\lfloor\left((\mathbf{S}_n\mathbf{V}_{i_j^u})^T \otimes (\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u})\right)\mathrm{vec}(\hat{\mathbf{H}}_{g,i_j^u})\right\rfloor_{n=1\ldots M_u}\right\rfloor_{k=1\ldots K_g^u} \\ \left\lfloor(\beta_B)^{\frac{1}{2}}\left\lfloor\left(\mathbf{V}_{i_j^u}^T \otimes \left((\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H)^T\mathbf{S}_n\right)\right)\mathrm{vec}(\hat{\mathbf{H}}_{g,i_j^u})\right\rfloor_{n=1\ldots M_B}\right\rfloor_{k=1\ldots K_g^u} \end{bmatrix}$$

$$\boldsymbol{\omega}_{4,gj} = \begin{bmatrix} \left\lfloor\vartheta_j^g\mathrm{vec}(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\hat{\mathbf{H}}_{g,j}\mathbf{V}_{i_j^d})\right\rfloor_{\substack{k=1\ldots K_g^u \\ i=1\ldots K_j^d}} \\ \left\lfloor(\kappa_B)^{\frac{1}{2}}\left\lfloor\left((\mathbf{S}_n\mathbf{V}_{i_j^d})^T \otimes (\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u})\right)\mathrm{vec}(\hat{\mathbf{H}}_{g,j})\right\rfloor_{n=1\ldots M_B}\right\rfloor_{\substack{k=1\ldots K_g^u \\ i=1\ldots K_j^d}} \\ \left\lfloor(\beta_B)^{\frac{1}{2}}\left\lfloor\left(\mathbf{V}_{i_j^d}^T \otimes \left((\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H)^T\mathbf{S}_n\right)\right)\mathrm{vec}(\hat{\mathbf{H}}_{g,j})\right\rfloor_{n=1\ldots M_B}\right\rfloor_{\substack{k=1\ldots K_g^u \\ i=1\ldots K_j^d}} \end{bmatrix}$$

$$\boldsymbol{\Omega}_{1k_g,i_j} = \begin{bmatrix} \left(\mathbf{V}_{i_j^u}^T \otimes \mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\right) \\ (\kappa_U)^{\frac{1}{2}}\left\lfloor(\mathbf{S}_n\mathbf{V}_{i_j^u})^T \otimes (\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d})\right\rfloor_{n=1\ldots M_u} \\ (\beta_U)^{\frac{1}{2}}\left\lfloor\mathbf{V}_{i_j^u}^T \otimes \left((\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H)^T\mathbf{S}_n\right)\right\rfloor_{n=1\ldots M_d} \end{bmatrix}$$

$$\boldsymbol{\Omega}_{2,k_g j} = \begin{bmatrix} \left\lfloor\left(\mathbf{V}_{i_j^d}^T \otimes \mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\right)\right\rfloor_{i=1\ldots K_j^d} \\ \left\lfloor(\kappa_B)^{\frac{1}{2}}\left\lfloor(\mathbf{S}_n\mathbf{V}_{i_j^d})^T \otimes \mathrm{vec}(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d})\right\rfloor_{n=1\ldots M_B}\right\rfloor_{i=1\ldots K_j^d} \\ \left\lfloor(\beta_U)^{\frac{1}{2}}\left\lfloor\mathbf{V}_{i_j^d}^T \otimes \left((\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H)^T\mathbf{S}_n\right)\right\rfloor_{n=1\ldots M_d}\right\rfloor_{i=1\ldots K_j^d} \end{bmatrix}$$

$$\boldsymbol{\Omega}_{3,g i_j} = \begin{bmatrix} \left\lfloor\left(\mathbf{V}_{i_j^u}^T \otimes \mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\right)\right\rfloor_{k=1\ldots K_g^u} \\ \left\lfloor(\kappa_U)^{\frac{1}{2}}\left\lfloor(\mathbf{S}_n\mathbf{V}_{i_j^u})^T \otimes (\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u})\right\rfloor_{n=1\ldots M_u}\right\rfloor_{k=1\ldots K_g^u} \\ \left\lfloor(\beta_B)^{\frac{1}{2}}\left\lfloor\mathbf{V}_{i_j^u}^T \otimes \left((\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H)^T\mathbf{S}_n\right)\right\rfloor_{n=1\ldots M_B}\right\rfloor_{k=1\ldots K_g^u} \end{bmatrix}$$

$$\boldsymbol{\Omega}_{4,gj} = \begin{bmatrix} \left\lfloor\left(\mathbf{V}_{i_j^d}^T \otimes \mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\right)\right\rfloor_{\substack{k=1\ldots K_g^u \\ i=1\ldots K_j^d}} \\ \left\lfloor(\kappa_B)^{\frac{1}{2}}\left\lfloor(\mathbf{S}_n\mathbf{V}_{i_j^d})^T \otimes (\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u})\right\rfloor_{n=1\ldots M_B}\right\rfloor_{\substack{k=1\ldots K_g^u \\ i=1\ldots K_j^d}} \\ \left\lfloor(\beta_B)^{\frac{1}{2}}\left\lfloor\mathbf{V}_{i_j^d}^T \otimes \left((\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H)^T\mathbf{S}_n\right)\right\rfloor_{n=1\ldots M_B}\right\rfloor_{\substack{k=1\ldots K_g^u \\ i=1\ldots K_j^d}} \end{bmatrix}$$

where $\mathbf{S}_n$ is a selection matrix consisting of all zeros except for the nth element along the diagonal which is equal to 1,

$$\delta_{i,j}^{k,g}=\begin{cases}1 & \textit{if } (k,g)=(i,j)\\ 0 & \textit{otherwise}\end{cases} \quad \textit{and} \quad \vartheta_j^g=\begin{cases}0 & \textit{if } g=j\\ 1 & \textit{otherwise}\end{cases}.$$

Proof. The problem formulation in (27) is based on finding an equivalent form for the inner minimization of (26). Note that $\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d})$ and $\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u})$ are given by (15) and (16) where $\Theta=1$ since we are dealing with imperfect CSI. Also the CSI error, Δ, appears in these terms when we replace $\mathbf{H}$ with $\hat{\mathbf{H}}+\Delta$ from (4). Next it can be noticed that the problem is separable over each occurrence of the different types of CSI error [30], i.e. we can separate the problem over $\Delta_{k_g^d,i_j^u}$, $\Delta_{k_g^d,j}$, Δ_{g,i_j^u} and $\Delta_{g,j}$, and focus on one of them at a time to obtain a more useful formulation.

Starting with $\Delta_{k_g^d,i_j^u}$, this will only appear in terms containing $\mathbf{H}_{k_g^d,i_j^u}$, since $\mathbf{H}_{k_g^d,i_j^u}=\hat{\mathbf{H}}_{k_g^d,i_j^u}+\Delta_{k_g^d,i_j^u}$. Therefore from the overall cost function of (26), from the perspective of each $\Delta_{k_g^d,i_j^u}$, we are only concerned with

$$\begin{aligned}
T_{1,k_g i_j} &= \mathrm{Tr}\left(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\\
&+\mathrm{Tr}\left(\kappa_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u}\,\mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\\
&+\mathrm{Tr}\left(\beta_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathrm{diag}(\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H)\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\\
&= \mathrm{Tr}\left(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\\
&+\sum_{n=1}^{M_u}\mathrm{Tr}\left(\kappa_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u}\,\mathbf{S_n}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{S_n}^H\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\\
&+\sum_{n=1}^{M_d}\mathrm{Tr}\left(\beta_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{S_n}\mathbf{H}_{k_g^d,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{k_g^d,i_j^u}^H\mathbf{S_n}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right).
\end{aligned}$$

Using $\mathrm{Tr}(\mathbf{X}\mathbf{X}^H)=\|\mathrm{vec}(\mathbf{X})\|^2$, $\mathrm{vec}(\mathbf{XYZ})=(\mathbf{Z}^T\otimes\mathbf{X})\mathrm{vec}(\mathbf{Y})$ and introducing slack variable $m_{1,k_g i_j}$, this can be expressed as

$$T_{1,k_g i_j}=\|\boldsymbol{\omega}_{1,k_g i_j}+\boldsymbol{\Omega}_{1,k_g i_j}\mathrm{vec}(\Delta_{k_g^d,i_j^u})\|^2\le m_{1,k_g i_j}. \quad (28)$$

Thus, the inner minimization in (26) from the perspective of each occurrence of $\Delta_{k_g^d,i_j^u}$ is given by

$$\begin{aligned}
&\max_{m} -m_{1,k_g i_j}\\
&\text{s.t. } \|\boldsymbol{\omega}_{1,k_g i_j}+\boldsymbol{\Omega}_{1,k_g i_j}\mathrm{vec}(\Delta_{k_g^d,i_j^u})\|^2\le m_{1,k_g i_j}\\
&\quad \forall\{\Delta_{k_g^d,i_j^u}:||\mathrm{vec}(\Delta_{k_g^d,i_j^u})||\le\varepsilon_{k_g^d,i_j^u}\}. \quad (29)
\end{aligned}$$

Next, representing the inequality in (28) as

$$-\left(\boldsymbol{\omega}_{1,k_g i_j}+\boldsymbol{\Omega}_{1,k_g i_j}\mathrm{vec}(\Delta_{k_g^d,i_j^u})\right)^H\mathbf{I}\left(\boldsymbol{\omega}_{1,k_g i_j}+\boldsymbol{\Omega}_{1,k_g i_j}\mathrm{vec}(\Delta_{k_g^d,i_j^u})\right)+m_{1,k_g i_j}\ge 0$$

we can apply the Schur Complement Lemma, to formulate the constraints of (29) as

$$\begin{bmatrix} m_{1,k_g i_j} & \boldsymbol{\omega}_{1,k_g i_j}^H\\ \boldsymbol{\omega}_{1,k_g i_j} & \mathbf{I}\end{bmatrix}+\begin{bmatrix}\mathbf{0} & \mathrm{vec}(\Delta_{k_g^d,i_j^u})^H\boldsymbol{\Omega}_{1,k_g i_j}^H\\ \boldsymbol{\Omega}_{1,k_g i_j}\mathrm{vec}(\Delta_{k_g^d,i_j^u}) & \mathbf{0}\end{bmatrix}\succeq\mathbf{0}.$$

Additionally, applying Lemma 1 from Appendix A with $\xi=\varepsilon_{k_g^d,i_j^u}$, $\mathbf{B}=[\mathbf{0}\ \boldsymbol{\Omega}_{1,k_g i_j}^H]$, $\mathbf{C}=[-1\ 0]$, $\mathbf{D}=\mathrm{vec}(\Delta_{k_g^d,i_j^u})$ and $\mathbf{A}=\begin{bmatrix} m_{1,k_g i_j} & \boldsymbol{\omega}_{1,k_g i_j}^H\\ \boldsymbol{\omega}_{1,k_g i_j} & \mathbf{I}\end{bmatrix}$ this can be further represented as

$$\lambda_{1,k_g i_j}\ge 0,\begin{bmatrix} m_{1,k_g i_j}-\lambda_{1,k_g i_j} & \boldsymbol{\omega}_{1,k_g i_j}^H & \mathbf{0}\\ \boldsymbol{\omega}_{1,k_g i_j} & \mathbf{I} & -\varepsilon_{k_g^d,i_j^u}\boldsymbol{\Omega}_{1,k_g i_j}\\ \mathbf{0} & -\varepsilon_{k_g^d,i_j^u}\boldsymbol{\Omega}_{1,k_g i_j}^H & \lambda_{1,k_g i_j}\mathbf{I}\end{bmatrix}\succeq\mathbf{0}. \quad (30)$$

Using the same separation of variables principle for $\Delta_{k_g^d,j}$, Δ_{g,i_j^u} and $\Delta_{g,j}$, for each of these CSI error terms we only need to focus on specific parts of the cost function given by (31), (32) and (33) respectively.

$$\begin{aligned}
T_{2,k_g j}=\sum_{i=1}^{K_j^d}\Bigg[&\mathrm{Tr}\left(\kappa_B\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j}\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\mathbf{H}_{k_g^d,j}^H\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\\
&+\mathrm{Tr}\left(\mathbf{B}_{k_g^d}(\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}-\delta_{i,j}^{k,g}\mathbf{I})(\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}-\delta_{i,j}^{k,g}\mathbf{I})^H\mathbf{B}_{k_g^d}^H\right)\\
&+\mathrm{Tr}\left(\beta_U\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathrm{diag}(\mathbf{H}_{k_g^d,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{k_g^d,j}^H)\mathbf{U}_{k_g^d}^H\mathbf{B}_{k_g^d}^H\right)\Bigg] \quad (31)
\end{aligned}$$

$$\begin{aligned}
T_{3,g i_j}=\sum_{k=1}^{K_g^u}\Bigg[&\mathrm{Tr}\left(\kappa_U\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u}\mathrm{diag}(\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H)\mathbf{H}_{g,i_j^u}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)\\
&+\mathrm{Tr}\left(\mathbf{B}_{k_g^u}(\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}-\delta_{i,j}^{k,g}\mathbf{I})(\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}-\delta_{i,j}^{k,g}\mathbf{I})\mathbf{B}_{k_g^u}^H\right)\\
&+\mathrm{Tr}\left(\beta_B\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathrm{diag}(\mathbf{H}_{g,i_j^u}\mathbf{V}_{i_j^u}\mathbf{V}_{i_j^u}^H\mathbf{H}_{g,i_j^u}^H)\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)\Bigg] \quad (32)
\end{aligned}$$

$$\begin{aligned}
T_{4,gj}=\sum_{k=1}^{K_g^u}\sum_{i=1}^{K_j^d}\Bigg[&\vartheta_j^g\mathrm{Tr}\left(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)\\
&+(1-\vartheta_j^g)\left(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\Delta_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\Delta_{g,j}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)\\
&+\mathrm{Tr}\left(\kappa_B\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,j}\mathrm{diag}(\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H)\mathbf{H}_{g,j}^H\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)\\
&+\mathrm{Tr}\left(\beta_B\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathrm{diag}(\mathbf{H}_{g,j}\mathbf{V}_{i_j^d}\mathbf{V}_{i_j^d}^H\mathbf{H}_{g,j}^H)\mathbf{U}_{k_g^u}^H\mathbf{B}_{k_g^u}^H\right)\Bigg] \quad (33)
\end{aligned}$$

Following a process similar to the one outlined for $\Delta_{k_g^d,i_j^u}$ we can introduce additional slack variables such that

$$\begin{aligned}
T_{2,k_g j}&=\|\boldsymbol{\omega}_{2,k_g j}+\boldsymbol{\Omega}_{2,k_g j}\mathrm{vec}(\Delta_{k_g^d,j})\|^2\le m_{2,k_g j}\\
T_{3,g i_j}&=\|\boldsymbol{\omega}_{3,g i_j}+\boldsymbol{\Omega}_{3,g i_j}\mathrm{vec}(\Delta_{g,i_j^u})\|^2\le m_{3,g i_j}\\
T_{4,gj}&=\|\boldsymbol{\omega}_{4,gj}+\boldsymbol{\Omega}_{4,gj}\mathrm{vec}(\Delta_{g,j})\|^2\le m_{4,gj}.
\end{aligned}$$

The introduction of these slack variables allows us to formulate the corresponding constraints in the form of (30). Additional details are not provided here since the process required for each of $\Delta_{k_g^d,j}$, Δ_{g,i_j^u} and $\Delta_{g,j}$ follows the one already outlined for $\Delta_{k_g^d,i_j^u}$. After going through this procedure, we can express the original cost function from (26) as a summation of the slack variables and some additional terms in order to obtain the final problem formulation in (27). □

Since problem (27) is not jointly convex in $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{B}$ we apply the alternating optimization approach in Algorithm 1 to solve it [30][4] . In Step 2, to compute $\mathbf{U}$ we fix $\mathbf{V}$ and $\mathbf{B}$ and solve the resulting semi-definite programming (SDP) problem. In Step 3 instead of finding $\mathbf{W}$, we now want to find

[4]Note that some additional minor reformulations are required when solving for $\mathbf{U}$ and $\mathbf{B}$. In particular, we introduce slack variables to handle terms of the form $\|\mathbf{BU}\|_F^2$ in a manner similar to the process applied in (28).

$$\mathbf{E}^S_{k^d_g} = \mathbf{U}_{k^d_g} \sum_{j=1}^{G} \sum_{i=1}^{K^d_j} \hat{\mathbf{H}}_{k^d_g,j} \mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j} \hat{\mathbf{H}}^H_{k^d_g,j} \mathbf{U}^H_{k^d_g} + \mathbf{U}_{k^d_g} \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \hat{\mathbf{H}}_{k^d_g,i^u_j} \mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j} \hat{\mathbf{H}}^H_{k^d_g,i^u_j} \mathbf{U}^H_{k^d_g} - \mathbf{U}_{k^d_g} \hat{\mathbf{H}}_{k^d_g,g} \mathbf{V}_{k^d_g} - \mathbf{V}^H_{k^d_g} \hat{\mathbf{H}}^H_{k^d_g,g} \mathbf{U}^H_{k^d_g}$$
$$+ (\sigma^2_U + f_{k^d_g}) \mathbf{U}_{k^d_g} \mathbf{U}^H_{k^d_g} + \mathbf{U}_{k^d_g} \hat{\mathbf{F}}_{k^d_g} \mathbf{U}^H_{k^d_g} + \mathbf{I} \tag{35}$$

$$\mathbf{E}^S_{k^u_g} = \mathbf{U}_{k^u_g} \sum_{\substack{j=1\\ j\neq g}}^{G} \sum_{i=1}^{K^d_j} \hat{\mathbf{H}}_{g,j} \mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j} \hat{\mathbf{H}}^H_{g,j} \mathbf{U}^H_{k^u_g} + \mathbf{U}_{k^u_g} \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \hat{\mathbf{H}}_{g,i^u_j} \mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j} \hat{\mathbf{H}}^H_{g,i^u_j} \mathbf{U}^H_{k^u_g} - \mathbf{U}_{k^u_g} \hat{\mathbf{H}}_{g,k^u_g} \mathbf{V}_{k^u_g} - \mathbf{V}^H_{k^u_g} \hat{\mathbf{H}}^H_{g,k^u_g} \mathbf{U}^H_{k^u_g}$$
$$+ (\sigma^2_B + f_g) \mathbf{U}_{k^u_g} \mathbf{U}^H_{k^u_g} + \mathbf{U}_{k^u_g} \hat{\mathbf{F}}_g \mathbf{U}^H_{k^u_g} + \mathbf{I} \tag{36}$$

$$\bar{\mathbf{U}}^S_{k^d_g} = \mathbf{V}^H_{k^d_g} \hat{\mathbf{H}}^H_{k^d_g,g} \left[\sum_{j=1}^{G} \sum_{i=1}^{K^d_j} \hat{\mathbf{H}}_{k^d_g,j} \mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j} \hat{\mathbf{H}}^H_{k^d_g,j} + \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \hat{\mathbf{H}}_{k^d_g,i^u_j} \mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j} \hat{\mathbf{H}}^H_{k^d_g,i^u_j} + \hat{\mathbf{F}}_{k^d_g} + (\sigma^2_U + f_{k^d_g}) \mathbf{I} \right]^{-1} \tag{37}$$

$$\bar{\mathbf{U}}^S_{k^u_g} = \mathbf{V}^H_{k^u_g} \hat{\mathbf{H}}^H_{g,k^u_g} \left[\sum_{\substack{j=1\\ j\neq g}}^{G} \sum_{i=1}^{K^d_j} \hat{\mathbf{H}}_{g,j} \mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j} \hat{\mathbf{H}}^H_{g,j} + \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \hat{\mathbf{H}}_{g,i^u_j} \mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j} \hat{\mathbf{H}}^H_{g,i^u_j} + \hat{\mathbf{F}}_g + (\sigma^2_B + f_g) \mathbf{I} \right]^{-1} \tag{38}$$

$\mathbf{B}$ where $\mathbf{W} = \mathbf{B}^H\mathbf{B}$. Therefore, after replacing terms of the form $\alpha \log_2 |(\ln 2/\alpha)\, \mathbf{B}^H\mathbf{B}|$ with $2\,\alpha \log_2 |(\ln 2/\alpha)^{\frac{1}{2}}\, \mathbf{B}|$, we fix $\mathbf{V}$ and $\mathbf{U}$ and solve the resulting determinant maximization (MAX-DET) problem [36]. Finally, in Step 4 to compute $\mathbf{V}$, we fix $\mathbf{U}$ and $\mathbf{B}$ and solve the resulting SDP problem. All problems may be solved using standard convex optimization solvers. Note that analogous to the algorithm from [30], the alternating maximization approach applied here to solve (26) converges. This follows because each step of the iterations leads to a monotonic increase of the objective function; since the objective function is upper bounded, convergence is guaranteed.

V. Robust design with stochastic error model

For the stochastic CSI error model, all nodes have access to $\hat{\mathbf{H}}$ instead of $\mathbf{H}$. Therefore instead of focusing on the actual achievable DL and UL rates, we consider their lower bounds $R^S_{k^d_g}$ and $R^S_{k^u_g}$, where channel estimation errors are treated as noise [37].

Starting with the DL, under Gaussian signaling, $R^S_{k^d_g} = \log_2 |\mathbf{I} + \hat{\mathbf{\Phi}}^{-1}_{k^d_g} \hat{\mathbf{H}}_{k^d_g,g} \mathbf{V}_{k^d_g} \mathbf{V}^H_{k^d_g} \hat{\mathbf{H}}^H_{k^d_g,g}|$ where $\hat{\mathbf{\Phi}}_{k^d_g}$ is defined as

$$\hat{\mathbf{\Phi}}_{k^d_g} = \sum_{\substack{j=1\\ (i,j\neq k,g)}}^{G} \sum_{i=1}^{K^d_j} \hat{\mathbf{H}}_{k^d_g,j} \mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j} \hat{\mathbf{H}}^H_{k^d_g,j} + \hat{\mathbf{F}}_{k^d_g}$$
$$+ \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \hat{\mathbf{H}}_{k^d_g,i^u_j} \mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j} \hat{\mathbf{H}}^H_{k^d_g,i^u_j} + (\sigma^2_U + f_{k^d_g}) \mathbf{I}\,.$$

Here $\hat{\mathbf{F}}_{k^d_g}$ is defined similarly to (7) but has all instances of $\hat{\mathbf{H}}$ replaced by $\mathbf{H}$. Additionally, $f_{k^d_g}$ reflects the effect of the stochastic imperfect CSI and is given by

$$f_{k^d_g} \approx \eta_{UB}(1 + \kappa_B + \beta_U) \sum_{j=1}^{G} \sum_{i=1}^{K^d_j} \mathrm{Tr}\left(\mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j}\right)$$
$$+ \eta_{UU}(1 + \kappa_U + \beta_U) \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \mathrm{Tr}\left(\mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j}\right).$$

For the UL, assuming Gaussian signaling, we have $R^S_{k^u_g} = \log_2 |\mathbf{I} + \hat{\mathbf{\Phi}}^{-1}_{k^u_g} \hat{\mathbf{H}}_{g,k^u_g} \mathbf{V}_{k^u_g} \mathbf{V}^H_{k^u_g} \hat{\mathbf{H}}^H_{g,k^u_g}|$, where $\hat{\mathbf{\Phi}}_{k^u_g}$ is given by

$$\hat{\mathbf{\Phi}}_{k^u_g} = \sum_{\substack{j=1\\ (j\neq g)}}^{G} \sum_{i=1}^{K^d_j} \hat{\mathbf{H}}_{g,j} \mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j} \hat{\mathbf{H}}^H_{g,j} + \hat{\mathbf{F}}_g$$
$$+ \sum_{\substack{j=1\\ (i,j\neq k,g)}}^{G} \sum_{i=1}^{K^u_j} \hat{\mathbf{H}}_{g,i^u_j} \mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j} \hat{\mathbf{H}}^H_{g,i^u_j} + (\sigma^2_B + f_g) \mathbf{I}\,.$$

Here $\hat{\mathbf{F}}_g$ is defined parallel to (10) with $\hat{\mathbf{H}}$ replaced by $\mathbf{H}$ and f_g is equivalent to

$$f_g \approx \eta_{BB}(1 + \kappa_B + \beta_B) \sum_{j=1}^{G} \sum_{i=1}^{K^d_j} \mathrm{Tr}\left(\mathbf{V}_{i^d_j} \mathbf{V}^H_{i^d_j}\right)$$
$$+ \eta_{BU}(1 + \kappa_U + \beta_B) \sum_{j=1}^{G} \sum_{i=1}^{K^u_j} \mathrm{Tr}\left(\mathbf{V}_{i^u_j} \mathbf{V}^H_{i^u_j}\right).$$

Therefore for the stochastic CSI error model, the WSR problem we want to solve is

$$\begin{aligned} \max_{\mathbf{V}} \; & \sum_{g=1}^{G} \sum_{k=1}^{K^d_g} \alpha_{k^d_g} R^S_{k^d_g} + \sum_{g=1}^{G} \sum_{k=1}^{K^u_g} \alpha_{k^u_g} R^S_{k^u_g} \\ \text{s.t.} \; & \mathrm{Tr}(\mathbf{V}_{k^u_g} \mathbf{V}^H_{k^u_g}) \leq P_U \;\; \forall\, k, g \\ & \sum_{k=1}^{K^d_g} \mathrm{Tr}(\mathbf{V}_{k^d_g} \mathbf{V}^H_{k^d_g}) \leq P_B \;\; \forall\, g\,. \end{aligned} \tag{34}$$

Similar to the perfect CSI case, we will solve this problem by transforming it into a WMMSE one. To obtain the MSE matrices, we start with $\mathbf{E}_{k^d_g} = \mathbb{E}[(\mathbf{U}_{k^d_g}\mathbf{y}_{k^d_g} - \mathbf{s}_{k^d_g})(\mathbf{U}_{k^d_g}\mathbf{y}_{k^d_g} - \mathbf{s}_{k^d_g})^H]$ and $\mathbf{E}_{k^u_g} = \mathbb{E}[(\mathbf{U}_{k^u_g}\mathbf{y}_g - \mathbf{s}_{k^u_g})(\mathbf{U}_{k^u_g}\mathbf{y}_g - \mathbf{s}_{k^u_g})^H]$ and replace $\mathbf{H}$ with $\hat{\mathbf{H}} + \Delta$ from (4). Taking the expectation over $\mathbf{s}$, $\mathbf{n}$ and $\mathbf{\Delta}$ under an independence assumption, we obtain $\mathbf{E}^S_{k^d_g}$ in (35) for the DL and $\mathbf{E}^S_{k^u_g}$ in (36) for the UL.

The optimal MMSE receivers can be obtained by solving $\bar{\mathbf{U}}^S_{k^d_g} = \arg\min_{\mathbf{U}} \mathrm{Tr}(\mathbf{E}^S_{k^d_g})$ and $\bar{\mathbf{U}}^S_{k^u_g} = \arg\min_{\mathbf{U}} \mathrm{Tr}(\mathbf{E}^S_{k^u_g})$, resulting in (37) and (38) respectively. Applying these MMSE receivers, the MSE matrices in (35) and (36) can respectively be expressed as $\bar{\mathbf{E}}^S_{k^d_g} = (\mathbf{I} + \mathbf{V}^H_{k^d_g}\; \hat{\mathbf{H}}^H_{k^d_g,g} \hat{\mathbf{\Phi}}^{-1}_{k^d_g} \hat{\mathbf{H}}_{k^d_g,g} \mathbf{V}_{k^d_g})^{-1}$ and $\bar{\mathbf{E}}^S_{k^u_g} = (\mathbf{I} + \hat{\mathbf{H}}_{g,k^u_g} \mathbf{V}_{k^u_g} \hat{\mathbf{\Phi}}^{-1}_{k^u_g} \mathbf{V}^H_{k^u_g} \hat{\mathbf{H}}^H_{g,k^u_g})^{-1}$. Similar to the

$$\mathbf{X}_g^S = \hat{\mathbf{X}}_g + \eta_{UB}(1+\kappa_B+\beta_U)\sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \mathrm{Tr}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^H\mathbf{W}_{i_j^d}\right) + \eta_{BB}(1+\kappa_B+\beta_B)\sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \mathrm{Tr}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^H\mathbf{W}_{i_j^u}\right) \quad (43)$$

$$\mathbf{X}_{k_g^u}^S = \hat{\mathbf{X}}_{k_g^u} + \eta_{UU}(1+\kappa_U+\beta_U)\sum_{j=1}^{G}\sum_{i=1}^{K_j^d} \mathrm{Tr}\left(\mathbf{U}_{i_j^d}\mathbf{U}_{i_j^d}^H\mathbf{W}_{i_j^d}\right) + \eta_{BU}(1+\kappa_U+\beta_B)\sum_{j=1}^{G}\sum_{i=1}^{K_j^u} \mathrm{Tr}\left(\mathbf{U}_{i_j^u}\mathbf{U}_{i_j^u}^H\mathbf{W}_{i_j^u}\right) \quad (44)$$

perfect CSI case, using an argument parallel to the one from [21]–[23], we obtain

$$R_{k_g^d}^S = \log_2\left|(\bar{\mathbf{E}}_{k_g^d}^S)^{-1}\right| \quad \text{and} \quad R_{k_g^u}^S = \log_2\left|(\bar{\mathbf{E}}_{k_g^u}^S)^{-1}\right| \; . \quad (39)$$

This rate to MSE relationship allows us to establish the following theorem.

Theorem 3. *The stochastic CSI error WSR problem in (34) is equivalent to the WMMSE problem in (40), such that the global optimal solution for the precoders of the two problems are identical.*

$$\begin{aligned}
\min_{\mathbf{U},\mathbf{W},\mathbf{V}} &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\left[\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}^S) - \alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{W}_{k_g^d}\right| - \frac{\alpha_{k_g^d}}{\ln 2}b_d\right] \\
&+\sum_{g=1}^{G}\sum_{k=1}^{K_g^u}\left[\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}^S) - \alpha_{k_g^u}\log_2\left|\frac{\ln 2}{\alpha_{k_g^u}}\mathbf{W}_{k_g^u}\right| - \frac{\alpha_{k_g^u}}{\ln 2}b_u\right] \\
\textit{s.t.} \quad & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \le P_U \quad \forall\, k,g \\
& \sum_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \le P_B \quad \forall\, g
\end{aligned} \quad (40)$$

Proof. The proof for this theorem is analogous to the perfect CSI argument used in the proof of Theorem 1 and the method in [21]–[23]; therefore to avoid repetition we have omitted it. Note that for the stochastic CSI error problem, the optimal $\mathbf{W}$ are

$$\bar{\mathbf{W}}_{k_g^d}^S = \frac{\alpha_{k_g^d}}{\ln 2}(\bar{\mathbf{E}}_{k_g^d}^S)^{-1} \text{ and } \bar{\mathbf{W}}_{k_g^u}^S = \frac{\alpha_{k_g^u}}{\ln 2}(\bar{\mathbf{E}}_{k_g^u}^S)^{-1} \; . \quad (41)$$

□

Since (40) is not jointly convex in $\mathbf{V}$, $\mathbf{U}$ and $\mathbf{W}$ the alternating optimization approach from Algorithm 1 is used to solve it. For Step 2 we use (37) and (38) to calculate the optimal receivers as $\mathbf{U}_{k_g^d} = \bar{\mathbf{U}}_{k_g^d}^S$ and $\mathbf{U}_{k_g^u} = \bar{\mathbf{U}}_{k_g^u}^S$. In Step 3 the weights are calculated as $\mathbf{W}_{k_g^d} = \bar{\mathbf{W}}_{k_g^d}^S$ and $\mathbf{W}_{k_g^u} = \bar{\mathbf{W}}_{k_g^u}^S$ using (41). The optimal precoders can be obtained similar to the perfect CSI case by using the Lagrangian method. Therefore in Step 4 we calculate $\mathbf{V}_{k_g^d} = \bar{\mathbf{V}}_{k_g^d}^S$ and $\mathbf{V}_{k_g^u} = \bar{\mathbf{V}}_{k_g^u}^S$ as

$$\begin{aligned}
\bar{\mathbf{V}}_{k_g^d}^S &= [\mathbf{X}_g^S + \mu_g^S\mathbf{I}]^{-1}\hat{\mathbf{H}}_{k_g^d,g}^H\mathbf{U}_{k_g^d}^H\mathbf{W}_{k_g^d} \\
\bar{\mathbf{V}}_{k_g^u}^S &= [\mathbf{X}_{k_g^u}^S + \lambda_{k_g^u}^S\mathbf{I}]^{-1}\hat{\mathbf{H}}_{g,k_g^u}^H\mathbf{U}_{k_g^u}^H\mathbf{W}_{k_g^u}
\end{aligned} \quad (42)$$

where $\mathbf{X}_{g^d}^S$ and $\mathbf{X}_{k_g^u}^S$ are defined in (43) and (44), and μ_g^S and $\lambda_{k_g^u}^S$ are the Lagrange multipliers. Here $\hat{\mathbf{X}}_g$ and $\hat{\mathbf{X}}_{k_g^u}$ are defined similar to $\mathbf{X}_g$ and $\mathbf{X}_{k_g^u}$ from (21) and (22) respectively but with $\mathbf{H}$ replaced by $\hat{\mathbf{H}}$.

Note that the convergence considerations in Remark 1 are also applicable to the alternating minimization approach applied to solve the stochastic CSI error problem (40).

VI. Extension to weighted DL rate maximization subject to a per UL user target rate

In addition to the total rate maximization design we also consider sum DL rate maximization subject to each UL user achieving a target rate of R_{UL}. The motivation behind this design is due to the fact that even if FD outperforms HD, this does not guarantee that all UL users are served evenly in every time slot. In some instances a UL user may achieve a lower rate in order to reduce the amount of interference present in the system. Therefore, we consider the following problem

$$\begin{aligned}
\max_{\mathbf{V}} \quad & \sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\alpha_{k_g^d}R_{k_g^d} \\
\text{s.t.} \quad & R_{k_g^u} \ge R_{UL} \quad \forall\, k,g \\
& \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \le P_U \quad \forall\, k,g \\
& \sum_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \le P_B \quad \forall\, g \; .
\end{aligned} \quad (45)$$

Theorem 4. *The WSR problem in (45) is equivalent to the WMMSE problem in (46), such that the global optimal solutions for the precoders of the two problems are identical.*

$$\begin{aligned}
\min_{\mathbf{U},\mathbf{W},\mathbf{V}} &\sum_{g=1}^{G}\sum_{k=1}^{K_g^d}\left[\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) - \alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{W}_{k_g^d}\right| - \frac{\alpha_{k_g^d}}{\ln 2}b_d\right] \\
\textit{s.t.} \quad & \left[\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) - \log_2\left|\ln 2\,\mathbf{W}_{k_g^u}\right| - \frac{b_u}{\ln 2}\right] \le -R_{UL} \quad \forall\, k,g \\
& \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \le P_U \quad \forall\, k,g \\
& \sum_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \le P_B \quad \forall\, g
\end{aligned} \quad (46)$$

Proof. Firstly it can be seen that the optimal $\mathbf{U}$ for (46) are the standard MMSE receivers $\bar{\mathbf{U}}_{k_g^d}$ and $\bar{\mathbf{U}}_{k_g^u}$ in (8) and (11) respectively. Secondly, fixing $\mathbf{U}$ and $\mathbf{V}$ and checking the first order optimality conditions for the weights we obtain their optimal values as

$$\bar{\mathbf{W}}_{k_g^d}^c = \frac{\alpha_{k_g^d}}{\ln 2}\bar{\mathbf{E}}_{k_g^d}^{-1} \text{ and } \bar{\mathbf{W}}_{k_g^u}^c = \frac{1}{\ln 2}\bar{\mathbf{E}}_{k_g^u}^{-1} \; . \quad (47)$$

Substituting for optimal $\mathbf{U}$ and $\mathbf{W}$ in (46) results in

$$\begin{aligned}
\min_{\mathbf{V}} & \sum_{g=1}^{G}\sum_{k=1}^{K_g^d} -\alpha_{k_g^d}\log_2\left|\bar{\mathbf{E}}_{k_g^d}^{-1}\right| \\
\text{s.t.} \quad & -\log_2\left|\bar{\mathbf{E}}_{k_g^u}^{-1}\right| \le -R_{UL} \quad \forall\, k,g \\
& \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \le P_U \quad \forall\, k,g \\
& \sum_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \le P_B \quad \forall\, g
\end{aligned} \quad (48)$$

which considering (9) and (12) is the same as (45). □

Since (46) is not jointly convex in $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{W}$ but is separately convex in each variable, it can be solved via alternating minimization. Having already obtained closed form

$$\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) = \|\boldsymbol{\phi}_{k_g^d}\|^2 = \left\| \begin{matrix} (\mathbf{I}\otimes\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,g})\mathrm{vec}(\mathbf{V}_{k_g^d}) - \mathrm{vec}(\mathbf{B}_{k_g^d}) \\ \lfloor(\mathbf{I}\otimes\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j})\mathrm{vec}(\mathbf{V}_{i_j^d})\rfloor_{\forall\, j=1\ldots G,i=1\ldots K_j^d,(i,j\neq k,g)} \\ \lfloor(\mathbf{I}\otimes\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u})\mathrm{vec}(\mathbf{V}_{i_j^u})\rfloor_{\forall\, j=1\ldots G,i=1\ldots K_j^u} \\ \kappa_B^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big(\mathrm{diag}(\mathbf{H}^H_{k_g^d,j}\mathbf{U}^H_{k_g^d}\mathbf{B}^H_{k_g^d}\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,j})\big)^{\frac{1}{2}}\Big)\mathrm{vec}(\mathbf{V}_{i_j^d})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^d}} \\ \kappa_U^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big(\mathrm{diag}(\mathbf{H}^H_{k_g^d,i_j^u}\mathbf{U}^H_{k_g^d}\mathbf{B}^H_{k_g^d}\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{H}_{k_g^d,i_j^u})\big)^{\frac{1}{2}}\Big)\mathrm{vec}(\mathbf{V}_{i_j^u})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^u}} \\ \beta_U^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big[\big(\mathrm{diag}(\mathbf{U}^H_{k_g^d}\mathbf{B}^H_{k_g^d}\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d})\big)^{\frac{1}{2}}\mathbf{H}_{k_g^d,j}\big]\Big)\mathrm{vec}(\mathbf{V}_{i_j^d})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^d}} \\ \beta_U^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big[\big(\mathrm{diag}(\mathbf{U}^H_{k_g^d}\mathbf{B}^H_{k_g^d}\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d})\big)^{\frac{1}{2}}\mathbf{H}_{k_g^d,i_j^u}\big]\Big)\mathrm{vec}(\mathbf{V}_{i_j^u})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^u}} \\ (\sigma_U^2+\beta_U\sigma_U^2)^{\frac{1}{2}}\,\mathrm{Tr}(\mathbf{B}_{k_g^d}\mathbf{U}_{k_g^d}\mathbf{U}^H_{k_g^d}\mathbf{B}^H_{k_g^d})^{\frac{1}{2}} \end{matrix} \right\|^2 \tag{50}$$

$$\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) = \|\boldsymbol{\phi}_{k_g^u}\|^2 = \left\| \begin{matrix} (\mathbf{I}\otimes\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,k_g^u})\mathrm{vec}(\mathbf{V}_{k_g^u}) - \mathrm{vec}(\mathbf{B}_{k_g^u}) \\ \lfloor(\mathbf{I}\otimes\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,j})\mathrm{vec}(\mathbf{V}_{i_j^d})\rfloor_{\forall\, j=1\ldots G,i=1\ldots K_g^d,(j\neq g)} \\ \lfloor(\mathbf{I}\otimes\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u})\mathrm{vec}(\mathbf{V}_{i_j^u})\rfloor_{\forall\, j=1\ldots G,i=1\ldots K_g^u,(i,j\neq k,g)} \\ \kappa_B^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big(\mathrm{diag}(\mathbf{H}^H_{g,j}\mathbf{U}^H_{k_g^u}\mathbf{B}^H_{k_g^u}\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,j})\big)^{\frac{1}{2}}\Big)\mathrm{vec}(\mathbf{V}_{i_j^d})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^d}} \\ \kappa_U^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big(\mathrm{diag}(\mathbf{H}^H_{g,i_j^u}\mathbf{U}^H_{k_g^u}\mathbf{B}^H_{k_g^u}\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{H}_{g,i_j^u})\big)^{\frac{1}{2}}\Big)\mathrm{vec}(\mathbf{V}_{i_j^u})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^u}} \\ \beta_B^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big[\big(\mathrm{diag}(\mathbf{U}^H_{k_g^u}\mathbf{B}^H_{k_g^u}\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u})\big)^{\frac{1}{2}}\mathbf{H}_{g,j}\big]\Big)\mathrm{vec}(\mathbf{V}_{i_j^d})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^d}} \\ \beta_B^{\frac{1}{2}}\lfloor\Big(\mathbf{I}\otimes\big[\big(\mathrm{diag}(\mathbf{U}^H_{k_g^u}\mathbf{B}^H_{k_g^u}\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u})\big)^{\frac{1}{2}}\mathbf{H}_{g,i_j^u}\big]\Big)\mathrm{vec}(\mathbf{V}_{i_j^u})\rfloor_{\substack{\forall\, j=1\ldots G\\ i=1\ldots K_j^u}} \\ (\sigma_B^2+\beta_B\sigma_B^2)^{\frac{1}{2}}\,\mathrm{Tr}(\mathbf{B}_{k_g^u}\mathbf{U}_{k_g^u}\mathbf{U}^H_{k_g^u}\mathbf{B}^H_{k_g^u})^{\frac{1}{2}} \end{matrix} \right\|^2 \tag{51}$$

expressions for optimal $\mathbf{U}$ and $\mathbf{W}$, we focus on obtaining $\mathbf{V}$. For fixed $\mathbf{U}$ and $\mathbf{W}$, we can express (46) as

$$\begin{aligned} \min_{\mathbf{V}} \quad & \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d}) \\ \text{s.t.} \quad & \mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u}) \le \Psi_{k_g^u} \ \ \forall\, k,g \\ & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \le P_U \ \ \forall\, k,g \\ & \sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \le P_B \ \ \forall\, g \end{aligned} \tag{49}$$

where $\Psi_{k_g^u} = -R_{UL} + \log_2|\ln 2\,\mathbf{W}_{k_g^u}| + b_u/\ln 2$.

Next using $\mathrm{Tr}(\mathbf{A}\mathbf{A}^H) = \|\mathrm{vec}(\mathbf{A})\|^2$ and $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{vec}(\mathbf{B})$ we can rewrite $\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d})$ and $\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u})$ as $\|\boldsymbol{\phi}_{k_g^d}\|^2$ in (50) and $\|\boldsymbol{\phi}_{k_g^u}\|^2$ in (51) respectively. This reformulation allows us to introduce slack variable t, such that $\|\boldsymbol{\phi}_{k_g^d}\|^2 \le t_{k_g^d}$, and cast (49) as the following problem

$$\begin{aligned} \min_{\mathbf{V},t} \quad & \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d} t_{k_g^d} \\ \text{s.t.} \ & \|\boldsymbol{\phi}_{k_g^d}\|^2 \le t_{k_g^d} \ \ \forall\, k,g \\ & \|\boldsymbol{\phi}_{k_g^u}\|^2 \le \Psi_{k_g^u} \ \ \forall\, k,g \\ & \mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H) \le P_U \ \ \forall\, k,g \\ & \sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H) \le P_B \ \ \forall\, g \end{aligned} \tag{52}$$

which after additional minor reformulations can be transformed into a second-order cone programming (SOCP) problem and then solved using standard convex optimization solvers.

Therefore to solve (46) we apply the alternating optimization process from Algorithm1. The optimal weights in Step 2 are calculated as $\mathbf{U}_{k_g^d} = \bar{\mathbf{U}}_{k_g^d}$ and $\mathbf{U}_{k_g^u} = \bar{\mathbf{U}}_{k_g^u}$ using (8) and (11). In Step 3, the optimal weights $\mathbf{W}_{k_g^d} = \bar{\mathbf{W}}^c_{k_g^d}$ and $\mathbf{W}_{k_g^u} = \bar{\mathbf{W}}^c_{k_g^u}$ are found using (47). In Step 4 the optimal precoders $\mathbf{V}_{k_g^d}$ and $\mathbf{V}_{k_g^u}$ are found by solving (52).

Proposition 1. *The alternating optimization process applied to solve (46) produces a convergent monotonically decreasing objective value sequence.*

Proof. See Appendix B. □

As a final remark, we would like to point out that it also possible to consider this problem under imperfect CSI, using the two models described in Section II-B. This will be considered as part of our future work.

VII. Simulation results

Our simulations follow the 3GPP LTE [16] specifications for multi-cell pico scenarios outlined in Table I. Channel gains between BSs and users, and between the users and the BSs themselves, are modeled as $\mathbf{H}_{r,t} = \sqrt{\varrho}\tilde{\mathbf{H}}_{r,t}$, where r represents the receiver, t represents the transmitter, $\tilde{\mathbf{H}}_{r,t}$ has elements distributed as $\mathcal{CN}(0,1)$ and $\varrho = 10^{-PL/10}$ with PL being the pathloss calculated according to Table I, depending on r and t. The SI channel, $\mathbf{H}_{g,g}$, is modeled as $\mathcal{CN}\big(\sqrt{K_H/(1+K_H)}\bar{\mathbf{H}}_{g,g}, (1/(1+K_H))\mathbf{I}_{M_B}\otimes\mathbf{I}_{M_B}\big)$ [1] where K_H is the Rician factor and $\bar{\mathbf{H}}_{g,g}$ is a deterministic matrix[5].

Throughout the simulations we fix $\alpha_{k_g^d} = \alpha_{k_g^u} = 1\ \forall k,g$ and $K_g^d = K_g^u = K\ \forall g$. We also set $\kappa_B = \kappa_U = \kappa$, $\beta_B = \beta_U = \beta$ and $\kappa = \beta$. Parameters κ and β jointly reflect the amount of transmitter and receiver distortion and, more importantly, κ on its own reflects the amount of residual SI at the FD BS as can be seen from (3). The larger the value, the larger both distortion and residual SI. Additionally for all algorithms we consider random precoder initialization and average the rate results in a Monte Carlo fashion over a number of randomly generated scenario realizations.

A. Perfect CSI results

Fig. 2 provides a comparison of the sum rates achieved by the FD beamformer design from Section III versus HD operation. For HD we consider the case where the BSs serve

[5] Without loss of generality, we set $K_H = 1$ and $\bar{\mathbf{H}}_{g,g}$ to be a matrix of all ones similar to [9], [20].

TABLE I: Parameter settings for simulations [16]

Parameter	Setting
Cell radius	40m
Bandwidth	10MHz
Thermal noise density	174dBm/Hz
Noise figure	BS: 13dB, user: 9dB
Max. transmit power	$P_B = 24$dBm, $P_U = 23$dBm
Min. distance	$D_{BS,BS-min} = 40$m $D_{BS,user-min} = 10$m
BS to BS pathloss (in dB, D in km)	LOS if $D < 2/3$: $98.4 + 20\log_{10}(D)$ LOS if $D \ge 2/3$: $101.9 + 40\log_{10}(D)$ NLOS: $101.9 + 40\log_{10}(D)$
BS to user pathloss (in dB, D in km)	LOS: $103.8 + 20.9\log_{10}(D)$ NLOS: $145.4 + 37.5\log_{10}(D)$
User to user pathloss (in dB, D in km)	if $D \le 50$m: $98.45 + 20\log_{10}(D)$ if $D > 50$m: $175.78 + 40\log_{10}(D)$
Shadowing standard deviation (in dB)	between BS & users, LOS: 3, NLOS: 4 between cells: 6
LOS probability (D in km)	$0.5 - \min(0.5, 5\exp(-0.156/D))$ $+ \min(0.5, 5\exp(-D/0.003))$

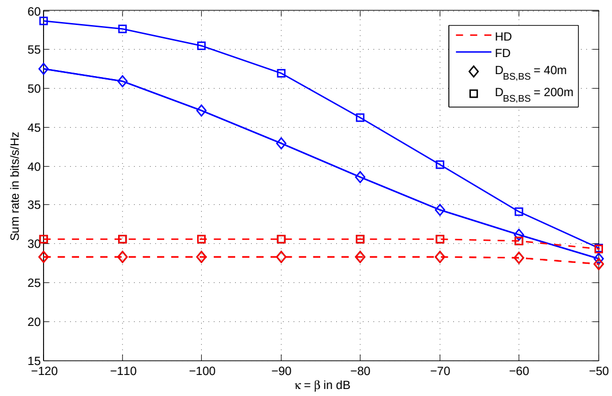


Fig. 2: Total sum rates achieved for scenario with $G = 2$, $K = 1$, $b_d = b_u = 1$, $M_B = 4$ and $M_d = M_u = 2$.

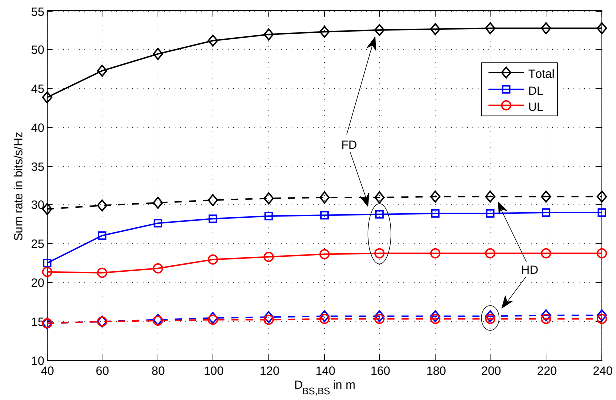


Fig. 4: Total sum rates achieved for scenario with $G = 2$, $K = 1$, $b_d = b_u = 1$, $M_B = 4$, $M_d = M_u = 2$ and $\kappa = \beta = -90$dB.

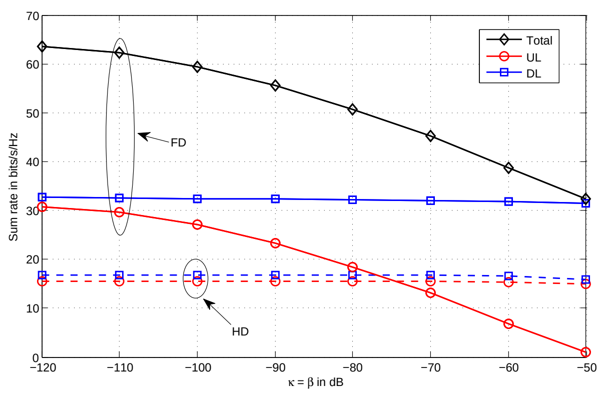


Fig. 3: Total sum rates achieved for scenario with $G = 2$, $K = 1$, $b_d = b_u = 1$, $M_B = 4$, $M_d = M_u = 4$ and $D_{BS,BS} = 100$m.

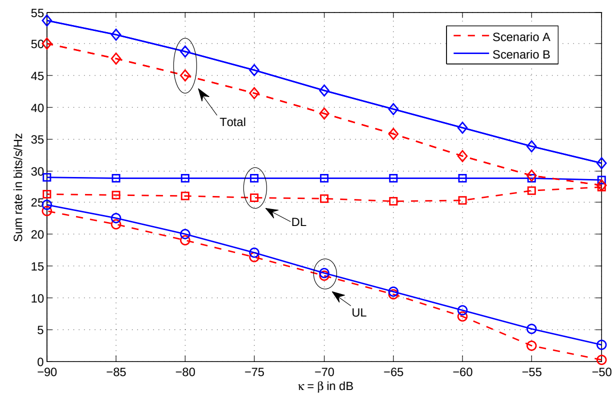


Fig. 5: Total sum rates achieved for scenario with $G = 2$, $K = 1$, $b_d = b_u = 1$, $M_B = 4$, $M_d = M_u = 2$ and $D_{BS,BS} = 100$m.

their corresponding DL and UL users separately in alternate channel uses, with the aim in each case being to maximize either the DL rate or the UL rate accordingly. As can be seen from Fig. 2 for $\kappa = \beta = -50$dB both HD and FD systems obtain similar rates, however FD outperforms HD for values of $\kappa = \beta < -50$dB. The amount of gain achieved varies with the $\kappa = \beta$ value and this is mainly due to the fact that the higher the transmitter distortion, κ, the more residual SI there is. This residual SI is a limiting factor for the UL rate which contributes a smaller portion of the total rate for larger κ. Such an effect can be seen more clearly in Fig. 3 which plots DL and UL rates separately.

From both Fig. 2 and Fig. 3, it can be noticed that as the value of $\kappa = \beta$ decreases the gain of FD over HD starts to increase significantly. In particular for Fig. 2 at $\kappa = \beta = -120$dB there is a gain of 1.92 for $D_{BS,BS} = 200$m and a gain of 1.85 for $D_{BS,BS} = 40$m. For FD the rate drop between achievable rates at $D_{BS,BS} = 200$m and at $D_{BS,BS} = 40$m is larger than the rate drop experienced by HD. This is due to the fact that when the BSs operate in FD there are more interference links than for HD, thus the negative impact of closer proximity between the cells affects FD more than HD.

The impact of inter-cell CCI on FD can be understood more clearly from Fig. 4, where we plot sum rate against the distance between BSs. As $D_{BS,BS}$ increases, inter-cell CCI decreases, thus the total achievable rate increases. An interesting effect can be noticed by looking at the separate FD DL and UL rate results in the range of 40m to 80m. The UL rate within this range remains approximately the same, however the DL rate has a significant increase. DL users experience inter-cell CCI from both BSs and UL users in other cells, thus a small increase in the distance between BSs contributes to a significant decrease in inter-cell CCI, allowing DL users to achieve higher rates. For this $D_{BS,BS}$ range, the BS to BS channel is very strong; implying that it is not advantageous in terms of the overall achievable rate to promote UL rate gain, hence the very small change in UL rate between 40m and 80m. A $D_{BS,BS}$ of around 100m or more is sufficient to overcome this issue, leading to a considerable increase in UL rate at 100m.

Having seen the effect of inter-cell CCI, next we investigate the effect of intra-cell CCI. In order to do so, we have devised two scenarios that fix the location of the BSs and the users, as shown in Fig. 6. For both scenarios A and B, the BSs are 100m apart and the distance between different cell DL and UL users is approximately 100m (100.50m for Scenario A and 100.32m for Scenario B), implying that the effect of inter-cell CCI is the same. However, the distance between same cell DL and UL users is only 10m for Scenario A and a much larger 56.49m for Scenario B. Fig. 5 provides some simulation results. As can be seen, scenario B achieves higher rates throughout; this is expected since Scenario B represents the lower interference case. Considering Scenario A and looking at the separate DL and UL rates it can be noticed that for example at $\kappa = \beta =$

-50dB the DL rate is around 27.5bits/s/Hz and the UL rate is nearly zero. For $\kappa = \beta = -50$dB the SI component is very high; this makes UL communication very difficult, thus DL communication is given priority. However as SI decreases, the UL rate starts to increase. This increase in UL rate in the high SI region comes at the expense of a slight decrease in the DL rate, due to the higher intra-cell CCI component. For scenario B, same cell UL and DL users are much further apart, thus the effect of intra-cell CCI is considerably reduced and this UL/DL rate trade-off does not occur.

B. Imperfect CSI results

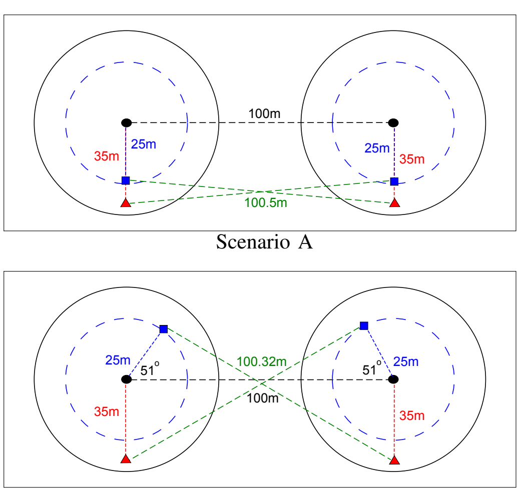
After establishing the gains of FD systems over HD ones, our next goal is to show how the FD imperfect CSI designs fare. Starting with the norm-bounded error design from Section IV we set $\varepsilon_{k_g^d, i_j^u} = \varepsilon_{k_g^d, j} = \varepsilon_{g, i_j^u} = \varepsilon_{g,j} = \varepsilon \; \forall k, g, i, j$ and obtain the results in Fig. 7. Note that channel strengths generated using the 3GPP LTE model from [16] are in the order of -30dB or lower, which is why for $\varepsilon = -30$dB achievable rates are close to zero. This also highlights why in the range of $\varepsilon = -30$dB to $\varepsilon = -35$dB, there is only a small difference in the rates achieved for different $\kappa = \beta$ values. Within this region the CSI error is considerably large, varying from being of the same order of magnitude as the strongest channels at -30dB to a third at -35dB; with CSI errors being so large, the error is more of a limiting factor on rate performance than transmitter and receiver distortion. The converse is true for lower ε regions. As the norm of the CSI error starts to decrease, the curves achieved for different $\kappa = \beta$ values become more distinct, indicating that distortion effects are more of a rate limiting factor than the CSI error. Naturally the curve for the lowest $\kappa = \beta$ settles at the highest rate value which is expected since this corresponds to the least amount of distortion and residual SI.

For the stochastic CSI error model, in Fig. 8 we plot the achievable rate for varying values of τ and ν, where $\tau = 0$ corresponds to perfect CSI for any ν. The robust design from Section V is compared with a naive version obtained by setting $\eta_{BB} = \eta_{BU} = \eta_{UB} = \eta_{UU} = 0$ in order to eliminate any robustness considerations. For fixed ν, rate decreases as τ increases; this is expected since larger τ values correspond to larger CSI errors. Additionally it can be noticed that the lower the ν the sharper is the rate decrease for varying τ, and the larger is the gain between the rate achieved by the robust beamformer versus the naive one. For $\nu = 1$, there is only a small difference between the performance of the robust and the naive designs, and the rate decrease for varying τ is also small. This behaviour is a reflection of the fact that previous studies with a similar error model show that $\nu = 1$ corresponds to perfect CSI from a degrees of freedom (DoF) perspective [35], [38].

Looking at both Fig. 7 and Fig. 8, it can be noticed that for both types of error the FD results deteriorate more than the HD ones for the same decrease in CSI quality. When using FD BSs there are more channel links between the various nodes than for the corresponding HD systems. Having an increased amount of links with imperfect knowledge results in a sharper rate decrease, thereby stressing the added importance of channel estimation and robust beamformer design for FD systems.

C. Results for target UL rate problem

For the problem from Section VI, which considers weighted DL rate maximization subject to a per UL user target rate, we



Scenario A

Scenario B

Fig. 6: Scenarios with same inter-cell CCI. Black circles represent the BSs, blue squares are UL users and red triangles are DL users.

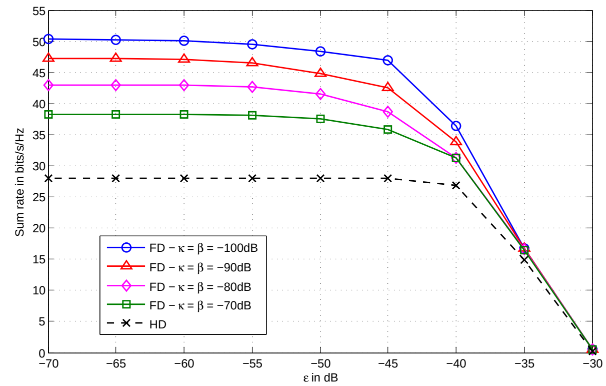


Fig. 7: Total sum rates achieved for different norm-bounded errors for scenario with $G = 2$, $K = 1$, $b_d = b_u = 1$, $M_B = 4$, $M_d = M_u = 2$ and $D_{BS,BS} = 100$m.

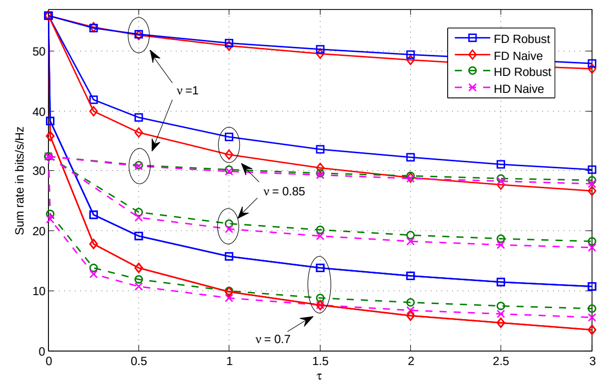


Fig. 8: Total sum rates achieved for different stochastic errors for scenario with $G = 2$, $K = 1$, $b_d = b_u = 1$, $M_B = 4$, $M_d = M_u = 4$, $D_{BS,BS} = 100$m and $\kappa = \beta = -90$dB.

have the results in Table II. This table provides a comparison between the DL rates achieved by the constraint design and the corresponding HD system which maximizes the total DL rate. Values written in the form $(a)^{*b\%}$ indicate that the problem is not always feasible for the considered target rate R_{UL}. Here $b\%$ represents the percentage of scenarios for which the problem was found to be feasible and a represents the average rate achieved over these feasible scenarios. UL rate results are not included, since provided that the chosen target is feasible a total UL rate of GKR_{UL} is achievable.

The gains of FD DL rate over HD DL rate range from 1.89 to 1.98 in Table II. On the other hand for the joint problem in Fig. 3, which considers the same system with $D_{BS,BS} = 100$m, there is a gain of 1.83 at $\kappa = \beta = -100$dB and 1.40 at $\kappa = \beta = -70$dB. Such a difference is mainly due to the fact that for FD κ and β are not only related to distortion but also to residual SI which makes UL communication more difficult. Constricting both FD and HD to achieve the same target UL rate removes the latter factor, thereby leading to higher gains over HD for the constraint problem as opposed to the joint one.

With respect to the feasibility of the chosen target rate, R_{UL}, it can be noticed that for a fixed R_{UL} the lower the distortion the more likely is the problem always feasible. For example at $D_{BS,BS} = 100$m and a target rate of $R_{UL} = 2.5$, feasibility goes from 12% at $\kappa = \beta = -70$dB to 100% at $\kappa = \beta = -90$dB. Such behaviour is expected because the higher the distortion, the stronger the SI and the more difficult it is to communicate in the UL. For the lowest distortion value of $\kappa = \beta = -100$dB, R_{UL} of up to around 8 is generally always feasible for $D_{BS,BS} = 100$m, this decreases to R_{UL} of up to around 5.5 for $D_{BS,BS} = 40$m. Naturally for $D_{BS,BS} = 100$m higher R_{UL} can be achieved than for $D_{BS,BS} = 40$, due to the stronger interference present in the latter scenario. This trend can also be seen by comparing the $D_{BS,BS} = 100$m and $D_{BS,BS} = 40$m results across Table II.

TABLE II: Sum DL rates achieved in bits/s/Hz for scenario with $G = 2$, $K = 1$, $b_u = b_d = 1$, $M_B = 4$ and $M_d = M_u = 4$.

$D_{BS,BS} = 40$m				
$\kappa = \beta$	FD			HD
(in dB)	$R_{UL} = 0.5$	$R_{UL} = 1.5$	$R_{UL} = 2.5$	
-100	32.01	31.96	31.87	16.67
-90	31.81	31.66	31.60	16.66
-80	31.60	31.50	$(31.44)^{*92\%}$	16.66
-70	$(31.38)^{*99\%}$	$(31.02)^{*48\%}$	$(30.48)^{*7\%}$	16.64

$D_{BS,BS} = 100$m				
$\kappa = \beta$	FD			HD
(in dB)	$R_{UL} = 0.5$	$R_{UL} = 1.5$	$R_{UL} = 2.5$	
-100	33.68	33.62	33.57	17.00
-90	33.50	33.42	33.40	16.98
-80	33.38	33.34	$(33.29)^{*98\%}$	16.98
-70	33.23	$(32.94)^{*64\%}$	$(33.28)^{*12\%}$	16.96

D. Convergence results

Finally, Fig. 9 illustrates the convergence behaviour of the proposed algorithms. For each algorithm we plot a randomly selected instance. In each case we set $\kappa = \beta = -90$dB and run for 30 iterations. For the perfect CSI problem we consider the system setup from Fig. 3. For the norm-bounded error problem we simulate the system from Fig. 7 with $\varepsilon = -45$dB. For the stochastic CSI error problem we consider the system from Fig. 8 with $\nu = 0.85$ and $\tau = 0.5$. For the constraint problem from Section VI we simulate the system in Table II at $D_{BS,BS} = 100$m with $R_{UL} = 1.5$. As can be seen all algorithms converge monotonically within a few steps.

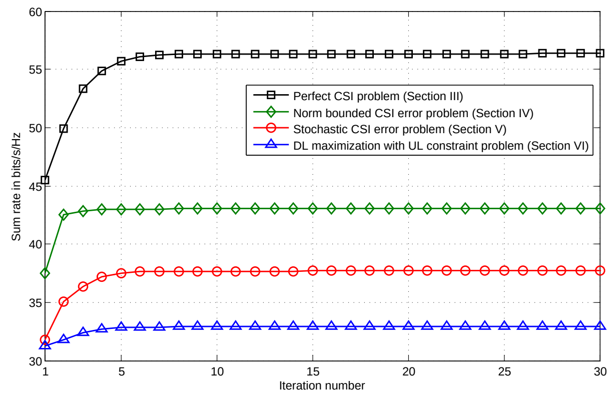


Fig. 9: Convergence behaviour of the proposed algorithms.

VIII. Implementation and complexity analysis

In order to simplify analysis, throughout this section we fix $M_B = M_d = M_u = M$, $K_d = K_u = K$ and $b_d = b_u = b$.

A. Implementation

All proposed algorithms can be implemented in a centralized manner, where a central processing site (CPS) collects all the required CSI, computes the required variables and then distributes them to the respective nodes. For this implementation a total of $M^2G^2(K^2 + 2K + 1)$ CSI elements need to be made available at the CPS to implement the algorithm. The CPS must then distribute the calculated precoders, resulting in $2GKMb$ matrix elements for all of $\mathbf{V}_{k_g^d}$ and $\mathbf{V}_{k_g^u}$.

Additionally the closed-form solution algorithms from Sections III and V may also be applied in a distributed manner. Similar to the implementations in [20], [22] and references therein, this requires all nodes to have knowledge of the channels directly linked to them, i.e. local CSI, and also assumes all receiving nodes can provide additional feedback information to transmitting nodes. Each receiving node locally estimates its interference-plus-noise covariance matrix, $\mathbf{\Phi}$. This metric is related to the MSE matrix which, when using an MMSE receiver, is given by $\bar{\mathbf{E}}_{k_g^d} = (\mathbf{I} + \mathbf{V}_{k_g^d}^H \mathbf{H}_{k_g^d,g}^H \mathbf{\Phi}_{k_g^d}^{-1} \mathbf{H}_{k_g^d,g} \mathbf{V}_{k_g^d})^{-1}$ for the DL and $\bar{\mathbf{E}}_{k_g^u} = (\mathbf{I} + \mathbf{V}_{k_g^u}^H \mathbf{H}_{g,k_g^u}^H \mathbf{\Phi}_{k_g^u}^{-1} \mathbf{H}_{g,k_g^u} \mathbf{V}_{k_g^u})^{-1}$ for the UL. Therefore $\mathbf{\Phi}$ can be used to calculate $\mathbf{U}$ and $\mathbf{W}$, which can then be made available to the transmitting nodes to calculate $\mathbf{V}$. Thus for a distributed implementation each node requires local CSI knowledge, resulting in a total of $2GKM^2$ elements across all users. Additionally $2GK(Mb + b^2)$ elements per iteration need to be fedback to the transmitting nodes to account for all of $\mathbf{U}$ and $\mathbf{W}$.

B. Complexity analysis

Starting with the closed-form solutions, we evaluate the order of the number of flops required to calculate the optimization variables using [39] which provides the number of flops required to perform standard mathematical operations. Taking (8) as an example, computing all receivers, $\mathbf{U}$, requires $\mathcal{O}\big(4G^2K^2(6M^3+2Mb^2)\big)$ flops for multiplications inside the inverse if $\mathbf{\Phi}$ is unavailable or $\mathcal{O}\big(2GK(4M^3+2Mb^2)\big)$ flops if $\mathbf{\Phi}$ is available, $\mathcal{O}(2GKM^3)$ flops for the inverse and $\mathcal{O}(8GKM^2b)$ flops for multiplying the inverse with the rest of the outside terms. To compute the weights, $\mathbf{W}$, we need to calculate the MSE matrix. The interference-plus-noise covariance matrix, $\mathbf{\Phi}$, is already available since it was previously used in the calculation of $\mathbf{U}$, therefore we only need $\mathcal{O}(2GKM^3)$ flops to calculate its inverse and $\mathcal{O}\big(2GK(4M^3+2M^2b+2Mb^2)\big)$ flops for multiplication. Taking (20) as an example, the calculation of each precoder, $\mathbf{V}$, requires $\mathcal{O}\big(4G^2K^2(6M^3+2M^2b+2Mb^2)\big)$ flops for multiplications inside the inverse to compute $\mathbf{X}$, $\mathcal{O}(2GKM^3)$ flops for the inverse and $\mathcal{O}\big(2GK(4M^2b+2Mb^2)\big)$ flops for multiplying the inverse with the rest of the outside terms.

For the norm-bounded error model we solve a number of SDP problems, the complexity of which is given by $\mathcal{O}(n^2\sum_i^I m_i^2)$ [40]. Here, n represents the total size of the variables being solved for and I is the total number of constraints, with each constraint i being of dimension m_i. In our case the complexity can be expressed as $\mathcal{O}\big((x_1+x_2)^2(z_1+z_2)\big)$, where $x_2=2G^2(K^2+2K+1)$ and $z_2=G^2K^2(1+M^2+b^2+2Mb^2)^2+2G^2K(1+M^2+Kb^2+2KMb^2)^2+G^2(1+M^2+K^2b^2+2MK^2b^2)^2+G^2(K^2+2K+1)^2$. When solving for $\mathbf{V}$, $x_1=2GKMb$ and $z_1=(GK+K)(Mb)^2$. When solving for $\mathbf{U}$, $x_1=2GKMb$ and $z_1=0$. When solving for $\mathbf{B}$ we have a MAX-DET problem. This is of higher complexity than SDP, however using the SDP complexity as a lower bound we have $x_1=2GKb^2$ and $z_1=0$.

In Section VI we solve an SOCP problem to obtain $\mathbf{V}$. The complexity of solving a general SOCP problem is given by $\mathcal{O}(n^2\sum_i^I m_i)$ [40], where the significance of the terms is the same as for the SDP complexity expression. Applying this to our problem we have $\mathcal{O}\big((2GKMb+GK)^2(GKMb+KMb+8G^2K^2Mb^3+4G^2K^2M^2b^2+2)\big)$.

IX. Conclusion

In this work we have addressed filter design for WSR maximization in multi-user multi-cell MIMO networks with FD BSs and HD users, taking into consideration CCI and transmitter and receiver distortion. Since WSR problems are non-convex, we have transformed them into WMMSE problems and proposed alternating optimization algorithms that are guaranteed to converge. Using the perfect CSI design as a starting point, we also consider robust beamformer design under two types of CSI error, namely norm-bounded error and stochastic CSI error. Simulation results for small cell scenarios show that replacing standard HD BSs with FD ones within this context can indeed increase achievable sum rate for low to intermediate distortion levels, and also confirm the robustness of the imperfect CSI designs. Additionally we also propose a DL rate maximization problem subject to each UL user achieving a desired target rate, which can be used in cases where it is important for each UL user to be equally served in every time slot.

Appendix A
Useful Lemma

Lemma 1. *[41] Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be given matrices, with $\mathbf{A}=\mathbf{A}^H$. Then, the relation*

$$\mathbf{A}\succeq\mathbf{B}^H\mathbf{D}\mathbf{C}+\mathbf{C}^H\mathbf{D}^H\mathbf{B}\quad\forall\,\mathbf{D}:\|\mathbf{D}\|\leq\xi$$

is valid if, and only if, there exists $\lambda\geq 0$ such that

$$\begin{bmatrix}\mathbf{A}-\lambda\mathbf{C}^H\mathbf{C} & -\xi\mathbf{B}^H\\ -\xi\mathbf{B} & \lambda\mathbf{I}\end{bmatrix}\succeq\mathbf{0}\,.$$

Appendix B
Proof of Propostion 1

Defining the following parameters

$$C_{k_g^u}(\mathbf{U},\mathbf{W},\mathbf{V})=\mathrm{Tr}(\mathbf{W}_{k_g^u}\mathbf{E}_{k_g^u})-\log_2|\ln 2\,\mathbf{W}_{k_g^u}|-\frac{b_u}{\ln 2}$$

$$C_{k_g^d}(\mathbf{U},\mathbf{W},\mathbf{V})=\mathrm{Tr}(\mathbf{W}_{k_g^d}\mathbf{E}_{k_g^d})-\alpha_{k_g^d}\log_2\left|\frac{\ln 2}{\alpha_{k_g^d}}\mathbf{W}_{k_g^d}\right|-\frac{\alpha_{k_g^d}}{\ln 2}b_d$$

we can express (46) as

$$\begin{aligned}\min_{\mathbf{U},\mathbf{W},\mathbf{V}}\ &\sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d}C_{k_g^d}(\mathbf{U},\mathbf{W},\mathbf{V})\\ \text{s.t.}\quad &C_{k_g^u}(\mathbf{U},\mathbf{W},\mathbf{V})\leq -R_{UL}\ \ \forall\,k,g\\ &\mathrm{Tr}(\mathbf{V}_{k_g^u}\mathbf{V}_{k_g^u}^H)\leq P_U\ \ \forall\,k,g\\ &\sum\nolimits_{k=1}^{K_g^d}\mathrm{Tr}(\mathbf{V}_{k_g^d}\mathbf{V}_{k_g^d}^H)\leq P_B\ \ \forall\,g\,.\end{aligned}\tag{53}$$

Assume that for (53) we have feasible solution $\{\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\}$ at the end of the (i)th iterate and feasible solution $\{\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i+1)\}$ at the end of the $(i+1)$th iterate. At the beginning of the $(i+1)$th iterate, to perform Step 2 of Algorithm 1, we fix the weights and precoders to $\mathbf{W}(i)$ and $\mathbf{V}(i)$ in order to obtain the updated receivers $\mathbf{U}(i+1)$. Since the updated receivers are MMSE ones, they are unique optimizers, therefore

$$C_{k_g^d}\big(\mathbf{U}(i+1),\mathbf{W}(i),\mathbf{V}(i)\big)\leq C_{k_g^d}\big(\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\big)\ \ \forall\,k,g\tag{54}$$

$$\begin{aligned}C_{k_g^u}\big(\mathbf{U}(i+1),\mathbf{W}(i),\mathbf{V}(i)\big)&\leq C_{k_g^u}\big(\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\big)\\ &\overset{(a)}{\leq}-R_{UL}\quad\forall\,k,g\end{aligned}\tag{55}$$

where (a) follows since $\{\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\}$ is feasible.

Next in Step 3, we fix the receivers and precoders to $\mathbf{U}(i+1)$ and $\mathbf{V}(i)$ in order to obtain the new weights $\mathbf{W}(i+1)$. The weights are updated using (47), which are unique optimizers, therefore

$$\begin{aligned}C_{k_g^d}\big(\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i)\big)&\leq C_{k_g^d}\big(\mathbf{U}(i+1),\mathbf{W}(i),\mathbf{V}(i)\big)\\ &\overset{(b)}{\leq}C_{k_g^d}\big(\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\big)\ \ \forall\,k,g\end{aligned}\tag{56}$$

$$
\begin{aligned}
C_{k_g^u}\big(\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i)\big) &\le C_{k_g^u}\big(\mathbf{U}(i+1),\mathbf{W}(i),\mathbf{V}(i)\big)\\
&\overset{(c)}{\le} C_{k_g^u}\big(\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\big)\\
&\le -R_{UL} \quad \forall\, k,g
\end{aligned}
$$

where (b) follows from (54) and (c) follows from (55).

At this stage we have intermediate solution $\{\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i)\}$ which is a feasible point, and the value of the cost function is given by

$$
\begin{aligned}
&\sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d} C_{k_g^d}\big(\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i)\big)\\
&\overset{(d)}{\le} \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d} C_{k_g^d}\big(\mathbf{U}(i),\mathbf{W}(i),\mathbf{V}(i)\big)
\end{aligned}
$$

where (d) follows from (56). Next, in Step 4, we fix the receivers and weights to $\mathbf{U}(i+1)$ and $\mathbf{W}(i+1)$ and solve (53) to obtain the new precoders $\mathbf{V}(i+1)$. Since with $\{\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i)\}$ the problem is known to be feasible, it follows that

$$
\begin{aligned}
&\sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d} C_{k_g^d}\big(\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i+1)\big)\\
&\le \sum\nolimits_{g=1}^{G}\sum\nolimits_{k=1}^{K_g^d} C_{k_g^d}\big(\mathbf{U}(i+1),\mathbf{W}(i+1),\mathbf{V}(i)\big)\ .
\end{aligned}
$$

As can be seen from the above process, the alternating optimization method applied to solve (46) produces a convergent monotonically decreasing objective value sequence, proving Proposition 1.

REFERENCES

[1] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of FD wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296-4307, Dec. 2012.

[2] J. I. Choi, M. Jain, K. Srinivasan, P. Levis, and S. Katti, "Achieving single channel, full duplex wireless communication," in *Proc. MOBICOM*, Sep. 2010, pp. 1-12.

[3] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proc. MOBICOM*, Sep. 2011, pp. 301-312.

[4] D. Bharadia and S. Katti, "Full duplex MIMO radios," in *Proc. USENIX*, Aug. 2014, pp. 359-372.

[5] A. Sahai, G. Patel, and A. Sabharwal, "Pushing the limits of full-duplex: design and real-time implementation," 2011. [Online]. Available: http://arxiv.org/abs/1107.0607

[6] H. Ju, E. Oh, and D. Hong, "Improving efficiency of resource usage in two-hop full duplex relay systems based on resource sharing and interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3933-3938, Aug. 2009.

[7] B. P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-Duplex MIMO Relaying: Achievable Rates Under Limited Dynamic Range," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 8, pp. 1541-1553, Sep. 2012.

[8] G. Zheng, I. Krikidis, and B. Ottersten, "Full-duplex cooperative cognitive radio with transmit imperfections," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2498-2511, May 2013.

[9] A. C. Cirik, R. Wang, Y. Rong, and Y. Hua, "MSE-Based Transceiver Designs for FD MIMO Cognitive Radios," *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2056-2070, Jun. 2015.

[10] D. Nguyen, L. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of Full-Duplex small cell wireless systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896-4910, Sep. 2014.

[11] S. Goyal *et al.*, "Improving small cell capacity with common-carrier full duplex radios," in *Proc. IEEE ICC*, Jun. 2014, pp. 4987-4993.

[12] S. Goyal, P. Liu, and S. Panwar, "User Selection and Power Allocation in Full Duplex Multi-Cell Networks," *IEEE Trans. Veh. Tech.*, vol.PP, no.99, pp.1-1.

[13] R. K. Mungara and A. Lozano, "Interference Surge in Full-Duplex Wireless Systems," in *Proc. IEEE Asilomar*, Nov. 2016, pp. 1-5.

[14] S. Goyal, P. Liu, S. Hua, and S. Panwar, "Analyzing a full-duplex cellular system," in *Proc. IEE CISS*, Mar. 2013, pp. 1-6.

[15] S. Wang, V. Venkateswaran, and X. Zhang, "Exploring full-duplex gains in multi-cell wireless networks: A spatial stochastic framework," in *Proc. IEEE INFOCOM*, Apr. 2015, pp. 855-863.

[16] 3GPP TR 36.828, "Further enhancements to LTE TDD for DL-UL interference management and traffic adaptation," Release 11, Jun. 2012.

[17] Y. S. Choi and H. Shirani-Mehr, "Simultaneous transmission and reception: Algorithm, design and system level performance," *IEEE Trans. Wireless Commun.*, vol. 12, no. 12, pp. 5992-6010, Dec. 2013.

[18] C. Nam, C. Joo, and S. Bahk, "Radio resource allocation with internode interference in full-duplex OFDMA networks," in *Proc. IEEE ICC*, pp. 3885-3890, Jun. 2015.

[19] D. Nguyen, L. N. Tran, P. Pirinen, and M. Latva-aho, "Precoding for Full Duplex Multiuser MIMO Systems: Spectral and Energy Efficiency Maximization," *IEEE Trans. Signal Process.*, vol. 61, no. 16, pp. 4038-4050, Aug.15, 2013.

[20] A. C. Cirik, R. Wang, Y. Hua, and M. Latva-aho, "Weighted Sum-Rate Maximization for FD MIMO Interference Channels," *IEEE Trans. Commun.*, vol. 63, no. 3, pp. 801-815, Mar. 2015.

[21] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792-4799, Dec. 2008.

[22] Q. Shi, M. Razaviyayn, and Z. Luo, "An Iteratively Weighted MMSE Approach to Distributed Sum- Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4331-4340, Sep. 2011.

[23] J. Shin and J. Moon, "Weighted-Sum-Rate-Maximizing Linear Transceiver Filters for the K-User MIMO Interference Channel," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 2776-2783, Oct. 2012.

[24] H. Ju, E. Oh, and D. Hong, "Improving efficiency of resource usage in two-hop full duplex relay systems based on resource sharing and interference cancellation," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 3933-3938, Aug. 2009.

[25] T. M. Kim, H. J. Yang, and A. J. Paulraj, "Distributed Sum-Rate Optimization for Full-Duplex MIMO System Under Limited Dynamic Range," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 555-558, Jun. 2013.

[26] W. Li, J. Lilleberg, and K. Rikkinen, "On Rate Region Analysis Of Half- and Full-Duplex OFDM Communication Links," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1688-1698, Sept. 2014.

[27] Q. Shi, M. Hong, X. Gao, E. Song, Y. Cai, and W. Xu, "Joint Source-Relay Design for Full Duplex MIMO AF Relay Systems," 2016 [Online]. Available: https://arxiv.org/abs/1607.03128

[28] Y. Sun, D. W. K. Ng, J. Zhu, and R. Schober, "Multi-Objective Optimization for Robust Power Efficient and Secure Full-Duplex Wireless Communication Systems," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 8, pp. 5511-5526, Aug. 2016.

[29] P. Ubaidulla and A. Chockalingam, "Relay Precoder Optimization in MIMO-Relay Networks With Imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5473-5484, Nov. 2011.

[30] J. Jose, N. Prasad, M. Khojastepour, and S. Rangarajan, "On Robust Weighted-Sum Rate Maximization in MIMO Interference Networks," in *Proc. IEEE ICC*, pp. 1-6, June 2011.

[31] N. Vučić, H. Boche, and S. Shi, "Robust Transceiver Optimization in Downlink Multiuser MIMO Systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576-3587, Sept. 2009.

[32] J. Wang, M. Bengtsson, B. Ottersten, and D. P. Palomar, "Robust MIMO Precoding for Several Classes of Channel Uncertainty," *IEEE Trans. Signal Process.*, vol. 61, no. 12, pp. 3056-3070, Jun. 2013.

[33] J. Wang and D. P. Palomar, "Robust MMSE Precoding in MIMO Channels With Pre-Fixed Receivers," *IEEE Trans. Signal Process.*, vol. 58, no. 11, pp. 5802-5818, Nov. 2010.

[34] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203-2214, May 2006.

[35] J. Maurer, J. Jaldén, D. Seethaler, and G. Matz, "Vector Perturbation Precoding Revisited," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 315-328, Jan. 2011.

[36] L. Vandenberghe, S.Boyd, and S. Wu, "Determinant Maximization with Linear Matrix Inequality Constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 2, pp. 499-533, 1998.

[37] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?," *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951-963, Apr. 2003.

[38] P. Aquilina and T. Ratnarajah, "Performance analysis of IA techniques in the MIMO IBC with imperfect CSI," *IEEE Trans. Commun.*, vol. 63, no. 4, pp. 1259-1270, April 2015.

[39] R. Hunger, “Floating Point Operations in Matrix-Vector Calculus,” 2007. [Online]. Available: https://mediatum.ub.tum.de/doc/625604/625604.pdf
[40] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, “Applications of second order cone programming,” *Linear Algebra APP.* vol. 284, no. 13, pp. 193-228, Nov. 1998.
[41] Y. Eldar and N. Merhav, “A competitive minimax approach to robust estimation of random parameters,” *IEEE Trans. Signal Process.*, vol. 52, no. 7, p. 1931-1946, Jun. 2004.

Paula Aquilina (S’14) received her B.Eng.(Hons.) degree in Electrical Engineering from the University of Malta in 2010 and an M.Sc. in Signal Processing and Communications from the University of Edinburgh in 2013. She is currently a Ph.D. candidate at the University of Edinburgh’s Institute for Digital Communications. Her main area of research is wireless communications, with particular focus on interference management and network information theory.

Ali Cagatay Cirik (S’13-M’14) received the B.S and M.S. degrees in telecommunications and electronics engineering from Sabanci University, Istanbul, Turkey, in 2007 and 2009, respectively, and Ph.D. degree in electrical engineering from University of California, Riverside in 2014. He held research fellow positions at Centre for Wireless Communications, Oulu, Finland and Institute for Digital Communications (IDCOM), University of Edinburgh, U.K. between June 2014 and November 2015. His industry experience includes internships at Mitsubishi Electric Research Labs (MERL), Cambridge, MA, in 2012 and at Broadcom Corporation, Irvine, CA, in 2013. Currently, he is working as a research scientist at Sierra Wireless, Richmond, Canada. He is also affiliated with University of British Columbia, Vancouver, Canada. His primary research interests are full-duplex communication, 5G non-orthogonal multiple-access (NOMA), MIMO signal processing, and convex optimization.

Tharmalingam Ratnarajah (A’96-M’05-SM’05) is currently with the Institute for Digital Communications, University of Edinburgh, Edinburgh, U.K., as a Professor in Digital Communications and Signal Processing and the Head of Institute for Digital Communications. His research interests include signal processing and information theoretic aspects of 5G and beyond wireless networks, full-duplex radio, mmWave communications, random matrices theory, interference alignment, statistical and array signal processing and quantum information theory. He has published over 300 publications in these areas and holds four U.S. patents. He is currently the coordinator of the FP7 project ADEL (3.7M€) in the area of licensed shared access for 5G wireless networks. Previously, he was the coordinator of the FP7 project HARP (3.2M€) in the area of highly distributed MIMO and FP7 Future and Emerging Technologies projects HIATUS (2.7M) in the area of interference alignment and CROWN (2.3M€) in the area of cognitive radio networks. Dr Ratnarajah is a Fellow of Higher Education Academy (FHEA), U.K., and an associate editor of the IEEE Transactions on Signal Processing.